

CUL-403337-2-0335076

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

2

With the advice and assistance of the Board of Editorial Consultants:

William Alston
Alan R. Anderson
Kurt Baier
Richard B. Brandt
Lewis W. Beck
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Michael Dummett

Peter Thomas Geach
Adolf Grünbaum
Carl G. Hempel
Jaakko Hintikka
Raymond Klibansky
Benson Mates
John A. Passmore
Günther Patzig
Richard H. Popkin

Wesley C. Salmon
George A. Schrader
Wilfrid Sellars
J. J. C. Smart
Wolfgang Stegmüller
Manley H. Thompson, Jr.
John Wild
G. H. von Wright
John W. Yolton

VOLUME 2/NUMBER 1

JANUARY 1965

CONTENTS

- | | | | |
|--|----|---|----|
| I. JONATHAN BENNETT: <i>Substance, Reality, and Primary Qualities.</i> | I | V. RICHARD G. HENSON: <i>What We Say</i> | 52 |
| II. RICHARD H. POPKIN: <i>The High Road to Pyrrhonism</i> | 18 | VI. PETER CAWS: <i>On Being in the Same Place at the Same Time</i> | 63 |
| III. G. NERLICH: <i>Presupposition and Entailment</i> | 33 | VII. JOHN J. FISHER: <i>Santayana on James: A Conflict of Views on Philosophy</i> . . | 67 |
| IV. R. F. HOLLAND: <i>The Miraculous</i> | 43 | VIII. A. MACC. ARMSTRONG: <i>Usage and Duty</i> | 74 |

UNIVERSITY OF PITTSBURGH PRESS

• AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes contributions by philosophers of any country on any aspect of philosophy, substantive or historical. However, self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

REPRINTS

Authors who are subscribers will receive 50 reprints gratis. Additional reprints can be purchased through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* for individual subscribers is six dollars and the price *per annum* for institutions is ten dollars. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues are sold at the rate of two dollars to individuals, and three dollars to institutions. Correspondence regarding subscriptions and back orders may be addressed directly to the publisher (University of Pittsburgh Press, University of Pittsburgh, Pittsburgh, Pennsylvania 15213).

* * *

The *American Philosophical Quarterly* is published quarterly in January, April, July, and October by the University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, Pennsylvania, 15213.
Second-class postage paid at Pittsburgh, Pennsylvania.

335076



AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston

Alan R. Anderson

Kurt Baier

Lewis W. Beck

Richard B. Brandt

Roderick M. Chisholm

L. Jonathan Cohen

James Collins

Michael Dummett

James M. Edie

Peter Thomas Geach

Adolf Grünbaum

Carl G. Hempel

Jaakko Hintikka

Raymond Klibansky

Benson Mates

John A. Passmore

Günther Patzig

Richard H. Popkin

Wesley C. Salmon

George A. Schrader

Wilfrid Sellars

J. J. C. Smart

Wolfgang Stegmüller

Manley H. Thompson, Jr.

G. H. von Wright

John W. Yolton



VOLUME 2 (1965)

UNIVERSITY OF PITTSBURGH PRESS

AMERICAN PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 2 (1965)

	Page
ACHINSTEIN, PETER <i>The Problem of Theoretical Terms</i>	193
ARMSTRONG, A. MACC. <i>Usage and Duty</i>	74
BARKER, STEPHEN F., <i>see</i> WESLEY C. SALMON, STEPHEN F. BARKER, HENRY E. KYBURG, JR.	
BENNETT, JONATHAN <i>Substance, Reality, and Primary Qualities</i>	1
BUTTS, ROBERT E. <i>Necessary Truth in Whewell's Theory of Science</i>	161
CAMPBELL, KEITH <i>Family Resemblance Predicates</i>	238
CAWS, PETER <i>On Being in the Same Place at the Same Time</i>	63
CHIHARA, C. S., AND J. A. FODOR <i>Operationalism and Ordinary Language</i>	281
COLEMAN, FRANCIS J. <i>Can a Smell or a Taste or a Touch be Beautiful?</i>	319
<i>Corrigenda</i>	325
DANTO, ARTHUR C. <i>Basic Actions</i>	141
DORE, CLEMENT <i>Seeming to See</i>	312
ELLIS, BRIAN <i>A Vindication of Scientific Inductive Practices</i>	296
FISHER, JOHN J. <i>Santayana on James: A Conflict of Views on Philosophy</i>	67
FODOR, J. A., <i>see</i> C. S. CHIHARA AND J. A. FODOR	
FRANKFURT, HARRY G. <i>Descartes' Validation of Reason</i>	149
GENTZEN, GERHARD <i>Investigations into Logical Deduction: II</i>	204
HENSON, RICHARD G. <i>What we Say</i>	52
HOLLAND, R. F. <i>The Miraculous</i>	43
KERFERD, G. B. <i>Recent Work on Presocratic Philosophy</i>	130
KHATCHADOURIAN, HAIG <i>Vagueness, Meaning, and Absurdity</i>	119
KYBURG, HENRY E., JR., <i>see</i> WESLEY C. SALMON, STEPHEN F. BARKER, HENRY E. KYBURG, JR.	
LLEWELYN, J. E. <i>Propositions as Answers</i>	305
McKINNON, ALASTAIR <i>Unfalsifiability and the Uses of Religious Language</i>	229
MACKIE, J. L. <i>Causes and Conditions</i>	245
MANDELBAUM, MAURICE <i>Family Resemblances and Generalization Concerning the Arts</i>	219
MARGOLIS, JOSEPH <i>Recent Work in Aesthetics</i>	182
MATTHEWS, GARETH B. <i>Augustine on Speaking from Memory</i>	157
NERLICH, G. <i>Presupposition and Entailment</i>	33
POPKIN, RICHARD H. <i>The High Road to Pyrrhonism</i>	18
SALMON, WESLEY C., STEPHEN F. BARKER, HENRY E. KYBURG, JR. <i>Symposium:</i> <i>Inductive Evidence</i>	265
SELLARS, WILFRID <i>Meditations Leibniziennes</i>	105
SHAFFER, JEROME A. <i>Recent Work on the Mind-Body Problem</i>	81

I. SUBSTANCE, REALITY, AND PRIMARY QUALITIES

JONATHAN BENNETT

TWO bad mistakes have been taken over from Berkeley by most philosophers who have read and assessed him with the casualness usually accorded to the great, dead philosophers. Each mistake is in the nature of a conflation or running together of two philosophical doctrines which ought to be kept apart, and thus a conflation also of the problems which the doctrines offer to solve. The doctrines in question are all expounded in Locke's *Essay Concerning Human Understanding*. They are: (1) a certain account of what it is for a property to be instantiated by something; (2) a certain account of the distinction between appearance and reality, or between how it is with me and how it is with the world; and (3) a thesis about primary and secondary qualities. Locke certainly accepted (2) and (3). His scathing attacks on (1) have usually been taken as a defence of it—here Locke has suffered the usual fate of the ironist.

In the first part of my paper I shall discuss the conflation by Berkeley, and by most English philosophers since his time, of (1) with (2). This conflation is, specifically, an *identification*: Berkeley and others have actually failed to see that (1) and (2) are distinct. The conflation of (2) with (3)—which I shall treat in the second part of the paper—has not usually taken the extreme form of an identifying of the two doctrines with one another. Occasionally, (3) is described as a "version" of (2), but a more common mistake is the milder one of treating (3) as being integrally connected with (2) in a way in which it is not.

In respect to both parts of the paper, I have been greatly helped by criticism from Peter Bell and Ian Hacking.

My interest in these conflations is philosophical rather than exegetical. If distinct false theories—such as (1) and (2)—are identified with one another, it will be harder to see why they are false and where the truth lies. Furthermore, I shall argue that there is something true and important which Locke, in his doctrine (3), was struggling to say about primary and secondary qualities. Yet his gestures in the right direction have not been followed up as they deserved; and it seems that

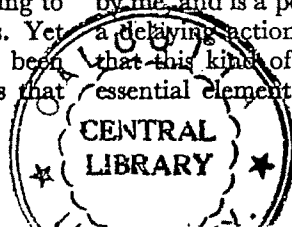
post-Lockean philosophers' neglect of the primary/secondary distinction has been due to their thinking that what Locke says about the distinction is an integral part, or one formulation, of a single monolithic doctrine of which (1) and (2) are also essential ingredients. This has tarred the primary/secondary distinction with the same brush as some things which have rightly been rejected, and so something important has been overlooked.

Through all the Berkeleian commentaries which identify (1) with (2), and wrongly connect (2) with (3), there is inevitably an appreciable haze of vagueness and failure of grasp. This does especial harm by nourishing the assumption that the study of philosophers like Locke and Berkeley is only a marginally useful activity which may be adequately conducted with the mind in neutral. In fact, it can be one of the most rewarding and demanding of philosophical exercises.

PART I

1. *The Substance Doctrine*

The account of property-instantiation which I call "Lockean," meaning that Locke said a good deal about it, is a view about the logic of subject-predicate statements. What concepts—or, as Locke would put it, what *ideas*—are involved in the subject of the statement that *The pen in my hand is expensive*? Certainly, the concepts of being a pen, and of being in my hand; but these are not enough, for the statement speaks of a *thing which* is a pen and is in my hand. What thing is this? I may answer that it is the purple thing which I now see before me; but when I say that the purple thing I now see is a pen and is in my hand, I speak of a *thing which* is purple, etc., and so my introduction of "purple" and "seen by me" still fails to capture the whole concept of the subject in the original statement. Even if I produce some non-trivial truth of the form "The . . . is purple, is seen by me, and is a pen in my hand," this can be only a delaying action. Sooner or later, I must admit that this kind of expansion is bound to omit an essential element from the concept of the pen in



my hand. What is missing is the concept of a *thing which . . .*: this is an ingredient in the concept of a *thing which is F* for each value of *F*, and is therefore not identical with the concept of a *thing which is F* for any value of *F*. This omnipresent constituent of any subject-concept is the concept of a property-bearer, or of a possible subject of predication. Let us call it the concept of a *substance*. It appears then that if any subject-predicate—or any existential—statements are true, there must be two basic sorts of item: (a) substances, and (b) qualities or properties. It is the special privilege of substances that they can bear or have or support qualities, and cannot in the same way be borne by anything else. We commit ourselves to the existence of “substances” in this sense every time we affirm of some property that it is instantiated by something or other: for a property to be instantiated is for there to be some substance which has or bears it.

I offer the foregoing paragraph as a rational reconstruction of one strand in the substantialism which Locke discusses in *Essay* II, xxiii, 1–4. In §2 he says: “The idea then we have, to which we give the general name substance, being nothing but the supposed but unknown support of those qualities we find existing, which, we imagine, cannot subsist *sine re substante*, without something to support them, we call that support *substantia*, which, according to the true import of the word is, in plain English, standing under, or upholding.” It is usual for Locke to say that we cannot “imagine” how qualities or accidents can exist unsupported, but the substantialism in question is certainly based, at least in part, upon logical considerations: some awareness of this is shown by Locke in II, xii, 4 and III, vi, 21, though the latter is not quite consistent.

Leibniz made a good remark about the Lockean theory of property-instantiation: “In distinguishing two things in [any] substance, the attributes or predicates, and the common subject of these predicates, it is no wonder that we can conceive nothing particular in this subject. It must be so, indeed, since we have already separated from it all the attributes in which we could conceive any detail” (*New Essays* II, xxiii, 2). This suggests, though it did not to Leibniz, the following argument. Suppose a substantialist were to say that any given item counts as a substance if and only if it has a certain property *S* which is definitive of substantiality. In that case, his account of what it is for a property to be instantiated, viz., that *P* is instantiated if and only if some substance bears

P, would say merely that *P* is instantiated if and only if some item is both *S* and *P*. His analysis of a statement about the instantiation of one property would thus yield, uselessly, a statement about the joint instantiation of two properties. A defender of the Lockean doctrine must therefore deny that substances are items of a certain kind: to be of a kind is to have the properties which define the kind, and the Lockean doctrine cannot allow that there are properties which substances must have in order to count as substances. But the claim that substances are items of a certain kind is the Lockean account of property-instantiation. The whole point and interest of the account lies in its claim that every subject-concept includes the concept of a certain kind of item whose special right and duty it is to bear properties.

The Lockean account must, therefore, be wrong. Its crucial error is the move from “There is a concept of a *thing which . . .*, which enters into every subject-concept” to “There is a kind of item about which nothing can be said except that items of that kind bear properties.” There are many kinds of things, but things do not form a kind. There is, perhaps, a “concept of a subject in general,” but it is to be elucidated in terms of the way in which more special concepts function in certain kinds of statement, and is not to be regarded as a concept which picks out a class of items.

2. The Veil-of-Perception Doctrine

Locke certainly did make a mistake about the distinction between what appears to be the case and what is really, or objectively, the case. His view is that the difference between seeing a tree, say, and being in a visual state as of seeing a tree though there is no tree to be seen, is the difference between having a sensory “idea” while in the presence of a real thing which is like the idea, and having such an idea while in the presence of no such thing. Sometimes he speaks only of a “correspondence,” “agreement,” or “conformity” between the sensory idea and the thing; and he also thinks that there is a causal relation between the two; but he speaks too of a “likeness,” and of ideas as “copies” of real things. This exposed him to a damaging attack from Berkeley who said that “An idea can be like nothing but an idea,” and that no sense attaches to the question whether human sensory states are informative of a real, objective world which is *like* them. This talk about ideas as like real things is associated with, and strongly

reinforces, Locke's mistaken handling of the question "Might it not be the case that there are no real things at all outside my mind? Can I be sure that the whole course of my experience is not just a dream?" Locke tries repeatedly to lay these sceptical doubts: see IV, ii, 14; iv, 4-5; and xi, 2-10. His arguments to this end are unsatisfactory, consisting as they do of *ad hominem* teasing of the sceptic and covert appeals to empirical evidence. Even opponents of phenomenism would now hesitate, I think, to follow Locke in his calm assumption that the question "Might it not be that there is no real extra-mental world?" requires an answer but stands in no need of criticism. Locke criticizes the moral character of the questioner, but his picture of the real world as represented by sensory states which "copy" it precludes his criticizing the question. This aspect of Locke's thought may be summed up in the remark that Locke puts the real world on the other side of the veil of perception, which explains my phrase "veil-of-perception doctrine." The word "doctrine" is misleading, though. Locke's treatment of the appearance/reality distinction is not prominent in the *Essay*: it appears mainly as a by-product of the mishandling of a certain sceptical question, and it has little of the weight or the deliberateness which go with a properly doctrinal status. That the veil-of-perception doctrine is usually credited to Locke as a *doctrine*—which he was consciously concerned to expound and defend—is due mainly to certain blunders which it is my present purpose to correct.

3. *The Two Doctrines in Berkeley*

The two philosophical views which I have sketched are distinct: one addresses itself to the question "What concepts do we use when we say *Something is F?*" while the other tackles the question "What is the difference between saying that *It is as though I were seeing a tree* and saying that *I see a tree?*" Although these are as different as chalk from cheese, they have been confidently identified with one another by Berkeley and by many other philosophers. Before explaining why Berkeley makes this remarkable mistake, I shall show that he makes it and in what ways.

Sometimes he does not make it at all, but treats one of the two Lockean doctrines in isolation from the other. In *Principles* § 49 he discusses the logical doctrine of substance without bringing in the veil-of-perception doctrine: "In this proposition,

A die is hard, extended and square, they will have it that the word *die* denotes a subject or substance, distinct from the hardness, extension, and figure which are predicated of it, and in which they exist. This I cannot comprehend: to me a die seems to be nothing distinct from those things which are termed its modes or accidents. And to say a die is hard, extended, and square is not to attribute those qualities to a subject distinct from and supporting them, but only an explication of the meaning of the word *die*." And in § 18-20, 86-88 there is a good part of the case against the veil-of-perception view, with no admixture of polemic against substance.

Nearly always, though, Berkeley welds the two doctrines together to form a single view about "material substance." Berkeley uses "matter" and its cognates to refer to Locke's purported "real things" which lie behind the veil of perception. (He also associates "matter" with Locke's views about primary qualities, but that raises issues which I shall discuss in Part II of this paper.) The word "substance," on the other hand, is especially associated with the Lockean account of property-instantiation. The phrase "material substance," then, which Berkeley uses lavishly and which hardly occurs in Locke, ensures that any discussion of one of the two doctrines has a good chance of becoming mixed up with a discussion of the other. Sometimes the mixture is fairly innocent: in *Principles* § 68, for example, Berkeley makes some shrewd remarks about substratum-substance, and, although he uses the word "matter" for what he is attacking, the attack itself is not seriously infected with anything which is appropriate to the veil-of-perception doctrine rather than the substance doctrine.

Often enough, however, the mixture is lethal. In *Principles*, § 16 Berkeley makes a point about substance, and not only refers to it as "matter" but also invokes "extension," which has nothing in particular to do with substratum-substance but does have to do with primary qualities and also with Locke's real world beyond the veil of perception: "It is said extension is a mode or accident of matter, and that matter is the substratum that supports it. Now I desire you that you would explain what is meant by matter's *supporting* extension. . . ."

Again, in § 17 Berkeley tries to locate the enemy: "If we inquire into what the most accurate philosophers declare themselves to mean by *material substance*, we shall find they acknowledge they have

no other meaning annexed to those sounds but the idea of being in general, together with the relative notion of its supporting accidents." This is a fair enough report of what Locke says not about "material substance" but about "substance." Berkeley adds that he does not understand the proffered account of the "meaning annexed to these sounds," and continues: "But why should we trouble ourselves any further in discussing this material *substratum* or support of figure and motion and other sensible qualities? Does it not suppose they have an existence without the mind? And is not this a direct repugnancy and altogether inconceivable?" He then launches off from "existence without the mind," etc., into an attack on the veil-of-perception doctrine! In this passage, a complaint against a wrong analysis of subject-concepts is jumbled together with a complaint against Locke's insufficiently idealist analysis of reality.

In § 37: "If the word *substance* be taken in the vulgar sense, for a combination of sensible qualities, such as extension, solidity, weight and the like: this we cannot be accused of taking away. But if it be taken in a philosophic sense, for the support of accidents or qualities without the mind; then indeed I acknowledge that we take it away, if one may be said to take away that which never had any existence, not even in the imagination." This might be taken to mean "Of course there are things which have properties, but in saying this we do not employ a concept of naked thinghood"; or it might be taken to mean "Of course there are real objects, but that statement can be analyzed purely in terms of mental states." There is no basis for preferring either interpretation.

In § 74: "But though it be allowed by the *materialists* themselves that matter was thought of only for the sake of supporting accidents. . . ."

Finally, in § 76: "If you stick to the notion of an unthinking substance, or support of extension, motion and other sensible qualities, then to me it is most evidently impossible there should be any such thing. Since it is a plain repugnancy that those qualities should exist in or be supported by an unperceiving substance."

These are some of the clearer expressions of the conflation; but the *Principles* and *Three Dialogues* contain many others which would suit my purpose even better if they did not also involve the further tangling of the two views so far discussed with Locke's view about primary qualities.

4. Why Berkeley Identified the Two Doctrines

This is not just a simple-minded blunder on Berkeley's part. His identification of the two Lockean doctrines flows naturally from his underlying assumption that the word "idea" can be used univocally to cover something in the nature of sensory states and something in the nature of concepts or meanings of words. This assumption enabled Berkeley to use "ideas of things" in such a way as to identify *qualities of things* with *sensory states which we have when we perceive things*. An idea of white for example is a certain kind of visual field; but it is also what I must be able to have in my mind if I am to understand the word "white," i.e., it is the meaning of the word "white," i.e., it is the property or quality of whiteness.

Some recent writers, sensing Wittgensteinian insights in Berkeley's theory of meaning, have denied that he takes "idea" in one of the two ways I have indicated. This, in my view, is a serious misreading of what Berkeley explicitly says about meaning and understanding; and it rides roughshod over the many passages in which he handles specific questions about meanings *solely* in terms of the possibility of bringing appropriate ideas into one's mind.

Since Berkeley uses "idea" in these two ways, it is natural that he should fail to distinguish the two Lockean doctrines; for each doctrine purports to offer an anchor for free-floating "ideas," one relating sensory states to the objectively real, and the other relating qualities to the things which have them. Furthermore, Berkeley can say of each Lockean doctrine that it over-populates the world: one by postulating "real things" which are logically dissociated from ideas (= sensory states), and the other by postulating "substances" which are something over and above collections of ideas (= qualities). The rather Berkeleian sentence "Things are just collections of ideas, not something over and above them" can be interpreted, taking ideas as qualities, as denying the Lockean account of property-instantiation; or, taking ideas as sensory states, as denying the veil-of-perception doctrine.

This diagnosis of the conflation is strongly confirmed in *Principles* § 78: "Qualities . . . are nothing else but *sensations* or *ideas*, which exist only in a mind perceiving them." Note also *Principles* § 9: "By matter therefore we are to understand an inert, senseless substance in which extension, figure, motion, do actually subsist, but it is evident from what we have already shown, that extension,

figure and motion are only ideas existing in the mind, and that an idea can be like nothing but another idea, and that consequently neither they nor their archetypes can exist in an unperceiving substance."

Special note should be taken of the phrase "sensible qualities," in which Berkeley often embodies his double use of "idea." For example in § 38: "But, say you, it sounds very ~~strange~~ ^{odd} to say we eat and drink ideas, and are clothed ~~with~~ ⁱⁿ ideas. I acknowledge it does so, the word *idea* ~~is~~ ^{being} used in common discourse to signify the several combinations of sensible qualities which are called *things* . . . But . . . the hardness or softness, the colour, taste, warmth and such like qualities which combined together constitute the several sorts of victuals and apparel, have been shown to exist only in the mind that perceives them; and this is all that is meant by calling them *ideas*. . ."

5. *The Two Doctrines in Locke*

The source of Berkeley's identification of the two doctrines is his double use of "idea"; but this he shares with Locke. Yet Locke does not run the substance doctrine together with the veil-of-perception doctrine: the two are distinct in Locke, as well as in fact. Since their non-distinctness does more or less follow from a premiss which Locke accepts—namely that "idea of *x*" can without ambiguity mean both "quality of *x*" and "appearance of *x*"—it must be conceded that Locke keeps the two doctrines apart only by betraying his basic premisses. This picture of Locke, as saying something true which he is committed to denying, is confirmed by certain details in the relevant parts of the *Essay*. These parts are not extensive. Contrary to the impression given by Berkeley, Locke does not have much to say about the substance which supports properties, or about the real world beyond the veil of perception: in the one case because he regards it as embarrassing and trivial and perhaps as just wrong, and in the other because he does not see that it involves an important mistake on a difficult philosophical problem. However, Locke's few discussions of substratum-substance and of related matters show that, although he has no intention of identifying the substance doctrine with the veil-of-perception doctrine, he cannot help expounding the former in words which would also be appropriate to the latter. In Locke's handling of the two doctrines, they drift together of their own accord,

(1) In the opening sections of *Essay* II, xxiii, Locke speaks of substance as something which we invoke when we become aware of "a certain number of simple ideas which go constantly together," or as something which is supposed to uphold "such combinations of simple ideas as are by experience and the observation of men's senses taken notice of to exist together." These expressions have to do with the instantiation of properties only if "idea" is taken to mean something like "property." But then, it seems, Locke is here raising not the general question "What is it for a property to be instantiated?" but the much more special question "What is it for a number of properties which go constantly together to be jointly instantiated?" This shift is bewildering; but it becomes intelligible if we remember that "ideas" may also be sensory states. For if we take "idea" to mean "sensory state," the phrases "ideas which go constantly together" and "combinations of simple ideas [which] exist together" may be taken to refer to certain kinds of dependable order in our experience. On that interpretation, the passages in question do not concern a queerly restricted version of the substratum-substance doctrine but rather concern the problems about objectivity or "reality" which are the province of the veil-of-perception doctrine. Locke makes no attempt to exploit these ambiguous phrases in order overtly to connect substance with what lies behind the veil-of-perception; but the basis for such a connection is there in the words he uses.

(2) In II, xxiii, 1, Locke says that when we note a number of sensory ideas going together, "not imagining how these simple ideas can subsist by themselves, we accustom ourselves to suppose some substratum wherein they do subsist and from which they do result; which, therefore, we call substance." Here again, substances are supposed to uphold "ideas"; and ideas must again be properties if the passage is to concern the substance doctrine at all. But, so construed, the passage says that substances are supposed to *cause* their own properties, and it is not clear why Locke should have thought that anyone believes that. Our puzzlement is removed if we remember that ideas can also be sensory states; for, on that reading of "from which [ideas] do result," it echoes that part of the veil-of-perception doctrine which says that real things cause our sensory states. As in the previous case, Locke here declines to cash in on this unhappy verbal overlap between the two doctrines in order explicitly to identify them with

one another. On the contrary, in the very next section he tries to drag "ideas" apart from "qualities," and thus to free the substance doctrine of any talk about causal relations by asserting that substances support qualities and that qualities cause sensory states: "If anyone will examine himself concerning his notion of pure substance in general, he will find he has no other idea of it at all, but only a supposition of he knows not what support of such qualities which are capable of producing simple ideas in us." If Locke had held firmly and consistently to that presentation of the matter, nearly every page of the *Essay* would have required revision. In fact, though, he is no more explicit or deliberate in his attempt to separate the two doctrines, and the two uses of "idea," than he is in allowing them to run together. Where Berkeley confidently identifies the two doctrines, Locke sometimes nearly identifies them and sometimes implicitly resists this identification; but the consequent tensions in his writing are not those of a man who has consciously located a problem.

(3) In II, xii, 4-6, Locke first distinguishes between "ideas of modes" and "ideas of substances." He clearly intends this to correspond to a distinction between adjectives and nouns, or between what may be said of a thing and things of which something may be said. This purely logical interpretation of the mode/substance distinction reappears at intervals throughout the *Essay*, for example in II, xiii, 19: "They who first ran into the notion of accidents, as a sort of real beings that needed something to inhere in, were forced to find out the word substance to support them," a remark which contains no hint of a restriction to substances of the special kind which Locke calls "real things." In the preceding section, too, Locke shows his awareness that the substance doctrine is supposed to account for property-instantiation generally, when he asks demurely whether God, finite minds, and bodies are all supposed to be "substances" in the same or different senses of the word. Yet even in his first introduction of "ideas of substances" and of the allegedly associated "supposed, or confused, idea of substance, such as it is," there is a dangerous reference to "distinct particular things subsisting by themselves." This last phrase could be taken to mean "things which exist independently of any perceiver," an interpretation which would connect "ideas of substances" with the veil-of-perception doctrine. Perhaps in that passage Locke is not taking "subsisting by themselves" in that way; but

he certainly does so later. In II, xxx, 4, he says, in effect, that in constructing complex ideas of modes we are subject only to the laws of logic: "There is nothing more required to those kinds of ideas to make them real, but that they be so framed, that there be a possibility of existing conformable to them." In the next section, however, he says that ideas of substances are subject to a more stringent requirement: "Our complex ideas of substances being made, all of them, in reference to things existing without us, and intended to be representations of substances as they really are, are no farther real than as they are such combinations of simple ideas as are really united and co-exist in things without us." This is a mistake: the propriety of a general noun no more depends upon its having instances than does the propriety of an adjective. My point, however, is that in making this mistake Locke very explicitly connects "ideas of substances" with questions about appearance and reality, and thus lays the foundation for connecting the latter with the doctrine of substratum-substance. I say only that he "lays the foundation" for this, because in this passage which so explicitly connects ideas of *substances* with "things without us" there is, interestingly, no mention at all of the idea of *substance*.

On this evidence, I think we may say that Locke did not wish to identify the two doctrines but was under pressure from his own presuppositions to do so. Had he thought to identify them he would, I think, have been deterred by the obvious absurdity of identifying the real world which may *resemble* our sensory ideas with the substratum-substances which may bear or uphold properties but which are in themselves *unqualified*. This manifest contradiction appears in most accounts of Locke—for example in those of Berkeley and Warnock—but I have yet to find a commentator who notices this feature of what is generally taken to be Locke's position.

6. Why Others Have Identified the Two Doctrines

So much for Berkeley and Locke; but what of those philosophers who have collapsed the substance doctrine into the veil-of-perception doctrine without having the excuse of an underlying mistake about the use of "idea"? What—one cannot help asking—do they think is happening in Berkeley's pages when they read the passages in which two or even three totally distinct questions are discussed at once? If they read Berkeley attentively

and critically, how do they think that his use of "material substance," etc., connects with Locke or with the truth? I cannot fully explain this propensity for taking Berkeley's problems at his own valuation of them; but the following hypothesis suggests how someone might come to accept the conflation without resting it directly on the double use of "idea."

One considers the distinction between appearance and reality and illustrates it by a situation in which one can say "It seems to me that I see something square, but is there really something square which I see?" One then puts this in the form: "I am in the presence of a manifestation, in my visual field, of squareness; but am I in the presence of something which is square?" The question whether what appears to be the case is really the case is thus quietly transmuted into the question whether a certain property has a possessor. One notes also that each question might—mistakenly but plausibly—be analyzed in terms of an elusive "something we know not what," and this further encourages one to believe that they are two versions of a single question of which Locke gave a single wrong analysis.

The train of thought indicated in my hypothesis is invalid. The question "Given that I seem to see something square, is there really something square which I see?" does not raise the question about property-instantiation which the Lockean doctrine of substance is supposed to elucidate. This is proved by a simple destructive dilemma.

(a) If we allow that my visual field contains a part which is square, then that part is the "thing which" is square, i.e., it bears the property of squareness with which I am confronted. It is a mistake to think that the Lockean concept of substance must be so handled that only physical or public or extra-mental objects are cases of substance-plus-properties. The whole point of the doctrine, as is often remarked even by those who perpetrate Berkeley's conflation, is that it separates the substance from *all* its properties and insists that for a property to be instantiated is for it to be borne by an item of which nothing can be said except that it bears that property. So: *if some part of my visual field is square, then I am not in the presence of a property for which I am seeking a bearer*, for the property in whose presence I am already has a bearer.

(b) If, more sensibly, we deny that anything in my visual field is itself square, and say only that

my visual field is similar to ones which I often have when I see something square, then my agnosticism about whether I see a square thing is agnosticism about whether I am in the presence of a manifestation of squareness at all. My question "Is the world at this point really as it appears to be?" is therefore not of the form "Is there a bearer for this property?" So: *if no part of my visual field is square, then I am not in the presence of a property for which I am seeking a bearer*, for I am not, in the required sense, "in the presence of a property" at all.

7. Some Examples

Here is one passage where Berkeley makes the shift I have described from "Does something real correspond to this sensory state?" to "Does something have this property?"

It is worth while to reflect a little on the motives which induced men to suppose the existence of material substance. . . . First, therefore, it was thought that colour, figure, motion, and the rest of the sensible qualities or accidents, did really exist without the mind; and for this reason it seemed needful to suppose some unthinking *substratum* or *substance* wherein they did exist, since they could not be conceived to exist by themselves. . . . (*Principles* § 73.)

O'Connor sees that there is a doctrine about substance of a purely logical kind. But he brings it in as an afterthought; dismisses it as an impossible interpretation of "the substratum theory," for no reason I can imagine except that he has taken Berkeley as his source for Locke; and shows, by his use of "something" in the first sentence, that he has not seen how distinct the two doctrines are:

It is certainly not logically necessary, or even true, that colours, for instance, cannot occur except as properties of a coloured something. If I stare at a light for a few seconds and then turn my gaze away, I shall see an "after-image" in the form of a coloured patch which certainly does not inhere in any substance. The supporter of the substratum theory of substance has either to claim (i) that the after-image is itself a substance or (ii) that it inheres in my visual field. (i) is a *reductio ad absurdum* of the substratum theory, though a sense datum would qualify as a substance in the logical sense of the word: it has properties without being itself a property of anything. . . .¹

With satisfying explicitness, Morris presents Berkeley's semi-phenomenalism as contradicting something said "largely on the credit of Aristotle's logic":

¹ D. J. O'Connor, *John Locke* (London, Penguin Books, 1952), pp. 80-81,

Berkeley has little difficulty in showing that the conception of material substance was in the philosophy of Locke no more than an uncriticized survival. Philosophers had always taken it for granted, largely on the credit of Aristotle's logic, that qualities must be supported by some underlying permanent self-subsistent substance. . . . Berkeley [argues against this] that throughout our whole experience of the physical world we never apprehend anything but sensible qualities and collections of sensible qualities. All we know of things or can know of them is what we perceive by sense; if there were more in things than this, we could not know it. This at once becomes clear, he says, if we consider what is meant by the term "exists". . . . "There was an odour, that is, it was smelled; there was a sound, that is to say, it was heard; a colour or figure, and it was perceived by sight or touch. This is all that I can understand by these and the like expressions." This doctrine is evidently based on the argument that whenever we are aware of a physical object, introspective analysis shows that there is nothing present in our mind but a number or collection of simple ideas of qualities; and it is taken by Berkeley to prove that knowledge simply consists in the awareness of sensible qualities.²

Warnock mixes the substance doctrine with the veil-of-perception doctrine by sliding smoothly from "matter" to "the essential 'support' of qualities":

We must seek to clarify Berkeley's disagreement with Locke about "matter" or "material substance." The central point of Berkeley's argument is that the expression "material substance" is *meaningless*, an empty noise. Locke, who held that our ideas are of qualities, had of course admitted that we do not "perceive" substance, that it is indeed "something we know not what"; but he thought that we must none the less accept this something, as being the essential "support" of qualities.³

On pp. 95-96 of his book Warnock tells us that according to Locke "there is a world of physical ('external') objects" which within certain limits "actually have the qualities which our ideas incline us to assign to them." Then on p. 109: "Locke had asserted the existence of 'matter', 'material substance', a *something* of which nothing could be either said or known." Does this introduce a second Lockean doctrine? Warnock seems not to think so. Is he then pointing out a flat inconsistency in a single Lockean doctrine? Apparently not: like Berkeley before him, Warnock presents as Locke's "doctrine" something which is flatly and

obviously inconsistent, yet does not call attention to this inconsistency because, one presumes, he has not noticed it. The double metaphor with which Warnock places the Lockean duplicate world "somehow behind or beneath" the world of experience reflects nicely his uncertainty as to just what view he wishes to attribute to Locke.

Examples can be found in nearly every extended discussion of Locke or Berkeley—in Fraser, Stephen, Huxley, Alexander, Hicks, Luce, Broad, Russell, Randall, Copleston—but not only in them. Ayer uses the phrase "sensible properties" in high Berkeleyan fashion to effect a slide from "the thing itself as opposed to anything which may be said about it" to "the thing itself [as opposed to] its appearances":

It happens to be the case that we cannot, in our language, refer to the sensible properties of a thing without introducing a word or phrase which appears to stand for the thing itself as opposed to anything which may be said about it. And, as a result of this, those who are infected with the primitive superstition that to every name a single real entity must correspond assume that it is necessary to distinguish logically between the thing itself and any, or all, of its sensible properties. And so they employ the term "substance" to refer to the thing itself. But from the fact that we happen to employ a single word to refer to a thing, and make that word the grammatical subject of the sentences in which we refer to the sensible appearances of the thing, it does not by any means follow that the thing itself is a "simple entity," or that it cannot be defined in terms of the totality of its appearances. It is true that in talking of "its" appearances we appear to distinguish the thing from the appearances, but that is simply an accident of linguistic usage. Logical analysis shows that what makes these "appearances" the "appearances of" the same thing is not their relationship to an entity other than themselves, but their relationship to one another.⁴

PART II

Locke distinguishes between "primary" and "secondary" qualities. A thing's primary qualities are its shape, size, spatial location, velocity, and degree of hardness; its secondary qualities are its color, temperature, smell, taste, and sound. Locke's attempt in *Essay* II, viii, 9, to give a general definition of this distinction is unsuccessful, but for present purposes the above lists suffice.

According to Locke, the secondary qualities of

² C. R. Morris, *Locke, Berkeley, Hume* (Oxford, The Clarendon Press, 1931), pp. 74-75.

³ G. J. Warnock, *Berkeley* (London, Penguin Books, 1953), p. 103.

⁴ A. J. Ayer, *Language, Truth and Logic* (London, Victor Gollancz, 1946), p. 42.

things are not "of" or "in" them in the same full-blooded sense as are their primary qualities. To say of something that it "is purple," for example, is to employ a natural and permissible *façon de parler*; while to say of something that it "is spherical" may be to state a plain fact in a way which requires neither gloss nor apology. I shall try to show that something true and interesting is misexpressed by this Lockean thesis, and that Berkeley's conflation of it with Locke's veil-of-perception doctrine reflects Berkeley's total failure to see what Locke was getting at in his discussion of primary and secondary qualities.

1. *The Phenol Argument*

Phenol-thio-urea tastes intensely bitter to 75 per cent of humans; to the rest it is tasteless. With a 25 per cent block of "non-tasters," we cannot say outright that the stuff is bitter: it tastes bitter to more people than not, but there is no such thing as "the" taste of it. If the non-tasters comprised only .001 per cent of all humans, then we could describe phenol-thio-urea as bitter without qualification: perhaps lemons are tasteless to .001 per cent of humans, but lemons are sour for all that. Suppose a world where phenol-thio-urea is unqualifiedly bitter, i.e., tastes so to almost everyone. Suppose further that a dynasty of world dictators begins intensive breeding of non-tasters and gradually allows the tasters to die out. This is good genetics: some of the tasters' offspring will be non-tasters, while the mating of two non-tasters can produce only non-tasting progeny. After a few dozen generations, phenol-thio-urea is tasteless to everyone living, so that there are as good grounds for calling phenol-thio-urea tasteless as for calling water tasteless.

This describes a course of events in which something (a) is bitter at one time, (b) is tasteless at a later time, and (c) does not itself change in the interim. This, on the face of it, is a contradiction; and we can resolve it only by saying that the stuff's bitterness is not one of "its properties" in the full-blooded sense in which a thing's losing one of its properties is its changing.

Similar arguments could be developed for the taste of any given kind of stuff, and also for colors, sounds, and smells. A simple genetic control would not always be available; but mass microsurgery might bring it about that no human could see any difference in color between grass and blood, and to do this would be to bring it about

that grass was the same color as blood. Similarly for other pairs of colors, and for tastes, sounds, and smells.

This is not one of those epistemological blockbusters which begins "Suppose that, as is logically possible, we were to wake up one morning and find that for some mysterious reason we were all . . . etc." The kind of story which I am telling is one in which, after the modification of the human frame has taken place, everyone knows just what has happened and how. Furthermore, the stories are more than just logically possible: we know how we could realize the tale about phenol-thio-urea, and the discovery of a surgical technique or a genetic control for any other discrimination which we make in respect of colors, tastes, sounds, or smells is scientifically well on the cards.

We may still call things green or sour or stinking or noisy, but philosophers should bear in mind the essentially relative nature of these adjectives and their like: "similar in color" means "looking similar in color to nearly everyone under normal conditions," and a careful metaphysic will take note of that fact.

The foregoing paragraphs contain what I shall for short call "the phenol argument." Before relating it to Locke and Berkeley, I should say at once that the argument is not valid. It depends upon the epitome which says that phenol-thio-urea is bitter at one time, tasteless at a later time, *and yet does not itself change in the interim*. The italicized clause is false: in the original story phenol-thio-urea does undergo a change, namely a change in respect of its taste. Admittedly it does not change its chemical structure, but to infer from this that it does not change at all is simply to beg the question in favor of primary qualities. The story shows that a thing may change in respect of its secondary qualities without changing its primary qualities; but this is not a contrast between primary and secondary qualities, for it is also true that a thing's primary qualities may change without any change in its secondary qualities. A fluid may have its primary qualities changed by the addition of a reagent, without changing in color, taste, or smell; and a knob of plasticine may be squashed flat.

It is natural to protest that *that* does not show the argument to be invalid, because squashing a piece of plasticine is doing something to it in a way in which the selective breeding of humans is not doing something to phenol-thio-urea. The phenol-thio-urea in the story, one might say, does not change in itself. I shall try below to show what

justice there is in these responses: to see the force of such phrases as "does not change in itself" is to see what the truth is about primary and secondary qualities.

2. *Locke and Berkeley on Secondary Qualities*

The phenol argument is mine, not Locke's: he does not suggest that a secondary quality of something might be altered by a species-wide physiological change. His discussions of primary and secondary qualities in II, viii, 9-26; xxiii, 11; and IV, iii, 11-13, 28-29 strongly suggest, however, that Locke would welcome the phenol argument as making his kind of point for his kind of reason. His own detailed arguments are more obviously unsatisfactory than the phenol argument; and yet even they give to some readers the impression that there is something true here which Locke is mishandling. I shall try to show that this impression is correct; but first let us see what Locke's arguments are, and what Berkeley does with them.

(1) Locke thinks that those of our sensory states which enable us to make secondary-quality discriminations between things can be explained in terms of the things' primary qualities: seen colors, for example, can be explained in terms of surface-textures, the impact upon our eyes of particles of light, and so on. But he stresses (II, viii, 11-13; IV, iii, 12-13, 28-29) that these explanations depend upon brute-fact, non-necessary, God-ordained correlations between our secondary-quality sensory states and the primary qualities which underlie and explain them. He seems to think—though he is reticent about this—that our seeings and feelings of the primary qualities of things have a necessary connection with the primary qualities themselves; perhaps because in that case there is supposed to be not only an explanatory or causal relationship but also a resemblance (II, viii, 15). Berkeley dismisses the talk about resemblances between ideas and bodies (*Principles* § 9); and argues that in any case there is only brute fact observed regularity in any of the connections ordinarily taken to be causal (§§ 25, 30-31). I assume that Berkeley is right on both these points, and that if Locke has got hold of a truth about primary and secondary qualities it must be sought elsewhere.

(2) In II, viii, 20, Locke says: "Pound an almond, and the clear white colour will be altered into a dirty one, and the sweet taste into an oily one. What real alteration can the beating of the

pestle make in any body, but an alteration of the texture of it?" Berkeley does not, I think, address himself directly to this; but he would have said that the beating of the pestle cannot make, or cause, any alteration whatsoever. His reasons for this lie outside my present scope, but this argument of Locke's is itself important, as I shall show. We may notice right away that the argument begs the question: Locke invites us to say that because the pestle can cause only primary-quality changes in the almond, the second-quality changes must therefore be primary-quality ones in disguise; but this can be rebutted by saying that beating something with a pestle can cause alterations other than primary-quality ones, as is proved by what happens to the color and taste of an almond when it is beaten with a pestle.

(3) In II, viii, 21, Locke points out that the same water may at once feel warm to one hand and cool to the other, which "figure never does, that never producing the idea of a square by one hand which has produced the idea of a globe by the other." Berkeley concedes the point about the warm/cool water phenomenon, but claims that it has primary-quality analogues, as can be discovered "by looking with one eye bare, and with the other through a microscope." (*First Dialogue*, pp. 219-222 in the Everyman edition).

(4) In II, xxiii, 11, Locke says that to the naked eye blood looks "all red," but through a good microscope it is seen as "some few globules of red swimming in a pellucid liquor; and how these red globules would appear, if glasses could be found that could yet magnify them 1,000 or 10,000 times more, is uncertain." Again, Berkeley agrees (*First Dialogue*, pp. 214-216) but says that analogous considerations apply to size, which is a primary quality (pp. 219-220).

(5) In II, viii, 16-18, Locke says that no reason can be given for saying that the heat is "actually in the fire" which would not also be a reason for saying that the pain is actually in the fire; yet it is clearly wrong to say the latter. Berkeley agrees with this cordially (*First Dialogue*, pp. 207-209), but takes it that here again there is no difference between primary and secondary qualities.

In each of (3), (4), and (5) Locke says something about secondary qualities which he thinks will show that they sit looser to the world than is usually thought; and in each case Berkeley kidnaps Locke's remark and uses it to prise primary qualities off the world as well. Yet Locke is *wrong*

in that part of each claim which Berkeley *accepts*. In (3), it does not follow—and should not even seem to follow—from the fact that we may err about temperatures that therefore things do not really have temperatures. In (4), the microscopic appearance of blood serves Locke's purpose only if it is possible that through a powerful enough microscope we should see the minute parts of blood as entirely colorless; and that is impossible since, for purposes of this argument, "colorless" must mean "invisible." In (5), it is not true that any grounds we could give for assigning temperatures to things would also be grounds for assigning pains to them.

Berkeley's view is not merely that what Locke says about secondary qualities is false unless it is so construed as to hold also for primary qualities, and that therefore Locke has failed to drive a wedge between the two sorts of quality. Berkeley genuinely agrees that things do not really have secondary qualities, and dissents only by saying that this is *also* true of primary qualities.

The explanation of this is as follows. Berkeley thinks that in agreeing with what Locke unclearly says about secondary qualities he is agreeing that a phenomenalist or idealist analysis ought to be given of statements about the secondary qualities of things, i.e., that talk about things' colors and smells and sounds, etc., is to be understood as shorthand for talk about certain sorts of sensory states. It is this thesis which he believes to hold also for talk about the primary qualities of things; it is the thesis which Berkeley offers as a rival to Locke's veil-of-perception account of the distinction between appearance and reality. Berkeley, in short, takes Locke's thesis about secondary qualities to be a *qualification* of his veil-of-perception doctrine. The latter says that there are facts about nonmental reality which are logically unconnected with facts about sensory states; and Berkeley takes the secondary-qualities doctrine to be an important rider to the effect that the genuinely extramental facts about reality are those which involve primary qualities only and do not include those which involve secondary qualities.

This account of what Berkeley is about explains the passages to which I have called attention as well as many more like them; and it is strongly confirmed by *Principles* § 14–15. If I am right about this, then Berkeley has completely misunderstood the kind of thing which Locke was trying to say; but I cannot justify this last claim without first saying what I think to be the truth about primary

and secondary qualities. Only in the light of what is true about primary and secondary qualities can we understand what Locke wanted to say about them.

3. *Color Blindness and Size "Blindness"*

Locke calls attention to the ways in which our perception of secondary qualities may vary according to the bodily condition of the percipient and according to the state of the percipient's environment. Berkeley rightly says that such variations also infect our perception of primary qualities but wrongly implies that the two sorts of quality are on a level in this respect. To see that they are not on a level, and why, is to grasp the truth after which Locke is fumbling.

I shall contrast two kinds of sensory aberration: in one, someone sees two things as being of the same color when in fact they are not, and in the other someone sees and feels two things as being of the same size when in fact they are not.

Suppose, then, that someone who is confronted by a red thing and a white thing convinces us that he sees them as having exactly the same color. He may believe us when we tell him that the things do have different colors; and if they differ in no other way we can, without asking him to trust us, prove to him that there is some difference between the two things which we see and he does not. Also, we may show him—or he may discover for himself—that the two objects differ in respect of the wave lengths of the light they reflect, and that wave lengths usually correlate with seen colors. But if he ignores other people's talk about the two objects, and ignores esoteric facts of optics, he may never discover that his seeing of the two objects as having the same color arises from a sensory defect in him. A failure of secondary-quality discrimination, in one who is otherwise sensorily normal, can—and sometimes does—persist unsuspected through any variations in distance or angle of view, light-conditions, mouth-washing, cold-curing, and so on.

Contrast this with the case of a size-"blind" man who, going by what he sees and feels, judges a certain drinking mug to have the same size as a certain cup, although in fact the former is both higher and wider than the latter. In such a case, we can place the cup inside the mug; or fill the mug with water, and then fill the cup from the mug and pour the remaining water on the ground; or place both vessels on a horizontal surface and draw the size-"blind" man's hand across the top

of the cup until it is stopped by the side of the mug; and so on. What are we to suppose happens when our size-"blind" man is confronted by these manipulations of the cup and the mug? There are just two relevant possibilities. (a) We may suppose that the size-"blind" man has a normal apprehension of what happens when we manipulate the cup and the mug, and therefore quite soon realizes that his original judgment about their sizes must have been mistaken. (b) We may suppose that in each case there is some supplementary inadequacy in his perception of what is done to the cup and the mug, or of the outcome of what is done, so that what he sees and feels still fits in smoothly with his original judgment that the two vessels are of the same size.

To adopt supposition (a) is just to admit that this case is radically different from that of color blindness. If the point of the latter were just that there are or could be aberrations in our perception of secondary qualities, then we could say the same of primary qualities; and we could add that it is absurd to deny that a certain kind of quality really is a quality of the things in the world, just because we do or might sometimes fail to discern it. What gives relevance and bite to color blindness, and to abnormality of secondary-quality perception generally, is the fact that any such abnormality can persist, not just for a few moments or under special conditions, without the victim's being given any clue to his abnormality by his other, normal sensory responses. The manipulations of the cup and the mug could be performed not by us but by the size-"blind" man himself: they involve ordinary commerce with familiar middle-sized objects, and are in a very different case from the color-blind man's attention to wave lengths or to other people's classifications of things by their colors.

If we want an analogy between size-"blindness" and color blindness, then, we must adopt supposition (b). But look at what this involves: the size-"blind" man must be unable to see or feel that the cup is inside the mug, or unable to see or feel that the mug has not momentarily stretched or the cup contracted; he must be unable to see or feel that the cup has been filled from the mug, or unable to see or feel that there is water left in the mug after the cup has been filled; he must be unable to see or feel that his hand is touching the cup as it moves across the top of it, or unable to see or feel his hand being stopped by the size of the mug. It will not do merely to suppose that as

the manipulations are performed he sees and feels *nothing*. To preserve the analogy with color blindness we must suppose that what he sees and feels gives him no reason for suspecting that there is something wrong with him; and so his visual and tactual states through all the manipulations of the cup and the mug must present no challenge to his belief that he is handling an ordinary pair of drinking vessels which are of the same size. This is bad enough, but there is worse to follow. If the size-"blind" man is to be unable to see or feel the water which remains in the mug after the cup has been filled from it, this will require yet further sensory aberrations on his part: if the water is poured over a lighted candle, or used to dissolve a lump of sugar, or thrown in the size-"blind" man's face, his perception of any of these events must also be appropriately abnormal if his original judgment is to remain unchallenged. Similarly with any of the other sensory aberrations with which we must prop up the initial one: each requires further props which demand yet others in their turn; and so on indefinitely.

The desired analogy with color blindness has collapsed yet again. In the case of color blindness, the sensory abnormality was not clued by the victim's other sensory responses although these were normal; but to keep the size-"blind" man in ignorance of his own initial sensory abnormality we have had to surround it with ever-widening circles of further abnormalities.

Strictly speaking, it is not quite correct to say that the single failure of color-discrimination could well remain unclued by the victim's other, normal sensory responses. For if he detects no difference in color between R_1 which is red and W_1 which is white, what are we to suppose that he makes of the color of a second red thing, R_2 , in relation to the color of W_1 ? If his only sensory aberration is to concern the comparison of R_1 with W_1 , then we must suppose that he sees no difference of color between R_2 and R_1 , and no difference of color between R_1 and W_1 , and yet sees a *large* difference of color between R_2 and W_1 . This is clearly unacceptable, and so we are in trouble here unless we suppose our man to be unable to see color differences between red things and white things generally. This, however, does not restore the analogy between color blindness and size "blindness." For the infectious spread of sensory aberrations around the single initial failure of size-discrimination does not involve merely other failures to discriminate sizes. The original aberration

tion can remain unclued only if it is backed by failures of shape-discrimination, movement-detection, sensitivity to heat, and so on. The single red/white failure spreads to red/white failures generally, but need spread no further; but the failure to discriminate sizes spreads endlessly into all the victim's perceptions of his environment.

As well as losing our analogy, we are also losing our grip on the initial datum of the size "blindness" case, namely that we can agree with the size-"blind" man about the identity of a certain cup and mug, disagreeing with him only about their relative sizes. For it has turned out that there are countless visible and tangible aspects of our environment in regard to which the size-"blind" man does not agree with us, so that it is now by no means clear that we can still assume that we share with him a sensory awareness of a single objective world.

4. *The Crucial Contrasts*

The foregoing discussion of color blindness and size "blindness" illustrates two crucial and closely related contrasts between primary and secondary qualities.

(1) There are countless familiar, exoteric, general facts about the connections between a thing's primary qualities and its ways of interacting with other things: a rigid thing cannot be enclosed within a smaller rigid thing; a thing cannot block another thing's fall to the earth without touching it; a cube cannot roll smoothly on a flat surface; a thing's imprint on soft wax matches the outline of the thing itself; and so on, indefinitely. The analogue of this does not hold for secondary qualities. Admittedly, there are connections between a thing's color, say, and its ways of behaving in relation to other things: in general, a brown apple will be more squashable than a green one; a blue flame will boil a pint of water faster than a yellow flame of the same size; a thing's color will correlate with the wave lengths of the light it reflects; and so on. But neither for colors nor for any other secondary qualities can we make, as we can for primary qualities, an enormously long list of obvious, familiar, inescapable connections of the relevant kind.

(2) Just because of the numerousness and familiarity of the connections between the primary qualities of things and their ways of interacting with one another, no clear sense attaches to the suggestion that something might persistently fail to obey these general connections. If a thing's pur-

ported size is belied by enough of its ways of interacting with other things, there is no point in saying that it does have that size. As against this, there would be a point in saying that a thing was red even if this were belied by the wave lengths of the light reflected from its surface, or by its flavor, hardness, chemical composition, etc. If in sunlight a given thing were indistinguishable in color from other things which were agreed to be red, then this fact could sensibly enough be reported in the words "That thing is red," even if we had to add one or more riders such as "... though its light-reflecting properties are atypical for red," or "... though its taste is atypical for red wine," or "... though its temperature is atypical for red iron." There is in fact a tight correlation between wave lengths of reflected light and the colors seen by most people in sunlight, and we therefore do not have to decide either for or against defining color-words in terms of how things look and treating the associated wave lengths as mere empirical correlates of colors. If the need for a decision did arise, however, we could choose to give our color terminology a purely visual basis and still have it doing pretty much the work which it does for us now. Analogous remarks apply also to all other secondary qualities. Not so, however, for primary qualities. As the discussion of size "blindness" showed, the interrelations between things in respect of their primary qualities are numerous and various and tightly interlocked. There seems to be little chance of inventing a partial breakdown of them such that those which survive the breakdown could form a basis for a working vocabulary of primary qualities. So far as I can see, the only kind of breakdown over which we could hope to keep control would be one involving the collapse of all but one of the normal correlates of some primary quality. For example, we might suppose a world in which things had reasonably reliable "sizes" if "the size of x " is defined solely in terms of the visual field presented by x when it is stationary and at some stated distance from the observer, and in which none of the other actual correlates of size continued to hold. This supposition does not, however, really provide us with a minimal sense of "size" analogous to the purely visual sense of "color"; for the supposition has cut away so much of what normally attends upon "size" that it almost certainly leaves no basis for a language of physical objects. It offers us a minimal sense of "size" while robbing us of everything which could have a size.

We can now see why the phenol argument is plausible, and why it really does show something about secondary qualities as against primary. We know what it would be like to be aware that the taste of phenol-thio-urea had been altered by means of a change of the human frame. But what could ever entitle us to say "Oranges, which used to be spherical, are now cubic; but this change has been brought about solely by a change in humans"? The difficulty here is not merely our ignorance of appropriate surgical or eugenic techniques, nor the scientific implausibility of suggesting that such techniques might be discovered: if the obstacles were only of that sort, then it would be historically impossible that the point brought out by the phenol argument should be one of which Locke was dimly aware. The trouble we meet in trying to reproduce a primary-quality analogue of the phenol argument is that we must either (a) allow the analogy to fail by supposing only that erstwhile spheres "look cubic" in some very restricted sense, e.g., in the sense of presenting visual fields like those now presented by sugar cubes when seen at rest (while in all other ways looking and feeling spherical); or (b) allow the analogy to fail by telling an astronomically complicated story in which not only the shapes of erstwhile spheres but also thousands of other aspects of the world were perceived differently; or (c) insulate shape from its present correlates by means of some radical conceptual revision which has no analogue in the phenol argument and which no one can see how to perform anyway. Of these alternatives, (a) and (b) do not produce the desired analogy, and perhaps (c) does not either; while (b) and (c) involve conceptual and empirical complications which we have no idea how to handle.

This difference between primary and secondary qualities is closely connected with the fact that the former alone involve the sense of touch. How they involve it, and what this has to do with the contrast I have drawn, are matters which I cannot go into here.

5. *The Relevance of This to Locke*

Locke says nothing about color blindness: it seems not to have been generally recognized in his day. Yet I maintain that the contrast I have drawn between primary and secondary qualities, and which I have approached through a discussion of sensory abnormalities, is one which Locke saw

dimly and was struggling to express and defend. My grounds for this contention are the following.

(1) The points which I have made could without absurdity be summed up in the Lockean remark that it is true of secondary qualities, in a way in which it is not of primary, that they are "merely" the powers which things have to affect us in certain ways.

(2) Locke was aware that primary qualities are all logically connected with solidity and extension, and these he regarded as definitive of "body" (II, iv, 1-2). Furthermore, he thought that the essentialness of "solid" and "extended" to "body" was connected with the different ways in which primary and secondary qualities are qualities "of" bodies, though he seems to have misunderstood the nature of the connection (II, viii, 9). My discussion indicates that Locke is right about the definition of "body," and right in his assumption that this is a deep conceptual fact which is not on a par with the dictionary definition of "brother" as "male sibling."

(3) Part of Locke's thesis about primary and secondary qualities is that if we knew enough we could give causal explanations, purely in primary-quality terms, for all our secondary-quality discriminations. Over the possibility of a purely primary-quality science, Locke had an optimism which was not at all justified by the state of physiology in the seventeenth century: note his calm assumption that *of course* the pestle's effect on the almond must be describable purely in primary-quality terms. My discussion of primary qualities shows why someone in Locke's position should so confidently assume that the final, perfect science will require only a primary-quality vocabulary.

(4) Several of Locke's examples share with the phenol example an emphasis upon the notion of a thing's changing in respect of a secondary quality without changing *in itself*. When Locke said that porphyry in the dark has no color, he erred; but he seems clearly to have had in mind the fact that our main basis—he would have said our only basis—for attributing colors to things is such that our color-reports can vary without any change in the things themselves. My discussion shows why this is plausible and how far it is true.

6. *Berkeley's Blunder*

I submit, then, that I have presented a truth about primary and secondary qualities and that it is this after which Locke was groping. Now, the

point which I have brought out has nothing to do with the veil-of-perception doctrine: it is not a version of that doctrine, or a qualification of or a rival to it. It operates on a different level altogether. One can state and explain what is interesting in the distinction between primary and secondary qualities—whether or not one goes so far as to say that secondary qualities are not “of” things as primary qualities are—only on the basis of normal assumptions about our entitlement to trust the evidence of our senses. What I have called Locke’s “veil-of-perception doctrine” is really just his mishandling of a certain sceptical question, and the latter makes sense only if it asks whether the objective world is, really, *in any way at all* as it appears to be. An affirmative answer must be given to this question before one can present the contrast between primary and secondary qualities.

To be fair, I must concede that the thesis about primary and secondary qualities can be taken as a qualification of the veil-of-perception doctrine in the following extremely minimal way. The veil-of-perception doctrine says that statements about objects are logically dissociated from statements about states of mind, and the primary/secondary thesis can be seen as conceding that this logical dissociation does not hold for statements attributing secondary qualities to objects but only for those attributing primary qualities to them. I think that it is because the relation between the two doctrines can be viewed like this that they are so often conflated; and it is therefore important to see what is wrong with this way of looking at the matter.

Considered as a qualification of the veil-of-perception doctrine, i.e., as a concession that not all statements about objects are logically dissociated from statements about states of mind, the primary/secondary thesis is just a bore. Even the most fervent super-Lockean would agree that some predicates of objects are connected with mental predicates; for example, that we commit ourselves to something about states of mind when we say that castor oil is nasty, that warm baths are soothing, that hair shirts are uncomfortable, or that the New York subway system is confusing. If Locke’s thesis about secondary qualities were important only as a concession that some predicates of objects are logically connected with mental predicates, it would be without any importance at all since it would be “conceding” what no one has ever denied. What makes it interesting is not its saying (a) “Some predicates of objects have

some logical connections with mental predicates,” but rather its saying (b) “Secondary-quality predicates of objects have *these* logical connections with mental predicates.” Now (b) does not offer any useful support to the view that there are logical connections between all predicates of objects and mental predicates: the Lockean view of the status of secondary qualities is no more a stage on the way to complete idealism or phenomenalism than is the Nazi valuation of Aryans a stage on the way to a belief in the worth and dignity of all men. In each case, the further step may consistently be taken; but in neither case is the taking of it just a further development of the line of thought by which the first stage was reached.

Just as the Lockean thesis about secondary qualities is not a significant *restriction* on the veil-of-perception doctrine, so the Lockean thesis about primary qualities is not—or need not be—a somewhat restricted *version* of the veil-of-perception doctrine. If to (a) “Some predicates of objects have *some* logical connections with mental predicates” we add the rider “but primary-quality predicates don’t,” then the result is indeed all of a piece with the veil-of-perception doctrine, and is thus in opposition to idealism or phenomenalism. But if to (b) “Secondary-quality predicates of objects have *these* logical connections with mental predicates” we add the rider “but primary qualities don’t,” the result says only that primary-quality predicates are not connected with mental predicates *in the way in which* secondary-quality predicates are. This presents no challenge at all to Berkeley or to any phenomenalist who knows what he is about: I have defended it myself, through my discussion of size “blindness,” without conceding a thing to the veil-of-perception doctrine.

This difference of level between the two theses is fairly clear in Locke’s own pages. In his battles with the sceptic, Locke does invoke empirical facts which are not legitimately available to him; but he does this covertly, and knows that he ought not to do it at all. As against this, his discussions of the primary/secondary contrast are riddled with open appeals to experimental evidence. (This is perfectly proper: a satisfactory treatment of the primary/secondary distinction must begin with empirical facts; though it ought, as Locke does not, to connect these with the relevant conceptual points.) Locke notes this explicitly in II, viii, 22: “I have, in what just goes before, been engaged in physical inquiries a little farther than perhaps I intended . . . I hope I shall be pardoned this

little excursion into natural philosophy, it being necessary, in our present inquiry, to distinguish the primary and real qualities of bodies, which are always in them . . . etc." Again, in IV, iii, 28, he denies that any "correspondence or connexion" can be found between our ideas of secondary qualities "and those primary qualities which (experience shows us) produce them in us." In the context of the battle against the sceptic, these references to "physical inquiries" and to what "experience shows us" would be merely grotesque. Still less do we find Locke mixing up the primary/secondary thesis with the question of substratum-substance. In his principal exposition of the former, II, viii, 9-26, the word "substance" does not occur.

As with the other conflation, so here there are some invitations to error, and thus some excuses for Berkeley, in Locke's pages. In particular, he says that our ideas of secondary qualities do not, while those of primary qualities do, resemble things themselves; and resemblance is also invoked in connection with the veil-of-perception doctrine. But the most that this shows is that Locke was unclear about the relation between the two theses: it could not show that they are—or even that Locke consciously thought them to be—related as Berkeley thinks they are.

(My attention was drawn by one of the referees for this *Quarterly*, and by Professor H. H. Price, to the relevance here of Locke's theory about real essences. A thing's "real essence" is that micro-physical primary-quality constitution of it which, according to Locke, is the causal basis for all its large-scale observable qualities, primary and secondary. Locke says that we know very little about real essences: "Though the familiar use of things about us takes off our wonder, yet it cures not our ignorance. When we come to examine the stones we tread on, or the iron we daily handle, we presently find that we know not their make; and can give no reason of the different qualities we find in them. . . . What is that texture of parts, that real essence, that makes lead and antimony fusible; and wood and stones not? What makes lead and iron malleable; antimony and stones not?" (III, vi, 9). What makes this important for Locke is just that since we do not *in fact* know much about real essences these cannot be the basis which we do now use for our classifications of physical things, and so the way is open for Locke to urge the candidacy of "nominal essences" as our actual basis for classification. But he sometimes (as in

III, iii, 17) gives the impression that he regards real essences as *necessarily* beyond the reach of our knowledge, and thus suggests that they have a "something we know not what" or a "beyond the veil" kind of status which smacks of both the substratum-substance and the veil-of-perception doctrines. I am sure that Locke did not hold as a considered opinion this strong view about the unknowability of real essences; but he undoubtedly does sometimes seem to hold it; and it is likely that Berkeley and others have in this way, as well as in the ways I have already discussed, been tempted to conflate the primary/secondary doctrine (of which the theory of real essences is an integral part) with one or both of the other two doctrines. This is the more probable since—as Professor Price has pointed out to me—the whole issue between Locke and Berkeley could be seen as a dispute between a proponent and an enemy of the microphysical approach to physical science. I find this last suggestion extremely illuminating—it captures Locke's picture of science as the minute *dissection* of large-scale objects, and Berkeley's picture of science as the intelligent *comparison* of large-scale objects with one another—but to follow it up here would take me too far afield.)

I conclude, then, that some things worth saying about the primary/secondary distinction are pointed to by Locke's discussion of it, and have no clear logical connection with the philosophical problem about the distinction between what appears to be the case and what is really the case.

It is for his insights into the latter problem that Berkeley is chiefly valued: he is rightly seen as a precursor of phenomenism, and even those who hold no brief for phenomenism agree that Berkeley taught us much about what goes wrong when the distinction between appearance and reality is divorced, as it is by Locke, from anything cashable in experience. But if we are to understand what is happening in Berkeley's pages, we must see through his appalling conflation of the question about the appearance/reality distinction with both the question about substance and that about the primary/secondary distinction. Consider for example the following passage from *Principles* § 9:

. . . they will have our ideas of the primary qualities to be patterns or images of things which exist without the mind, in an unthinking substance they call *matter*. By matter, therefore, we are to understand an inert, senseless substance, in which extension, figure, motion and so forth do actually subsist, but it is

evident from what we have already shown that extension, figure and motion are only ideas existing in the mind, and that an idea can be like nothing but another idea, and that consequently neither they nor their archetypes can exist in an unperceiving substance. Hence it is plain, that the very notion of what is called *matter*, or *corporeal substance*, involves a contradiction in it.

How can such a farrago as this be understood—how could anyone spell out in plain terms what it is that is being opposed here—except on the basis of an elaborate exposure of the two conflations? Many passages in *Principles* and *Three Dialogues* are similarly unintelligible until the two conflations have been understood and rejected; and Berkeley's writings are full of tensions which can be resolved only on that same basis. For example, according to the primary/secondary doctrine, things do have primary qualities in a way in which they do not have secondary; according to the veil-of-perception doctrine things may really have none of the properties we attribute to them; and of course substratum-substances cannot have, *qua* substances, any properties at all. And so, although "by matter, therefore, we are to understand an inert, senseless substance, in which extension, figure, motion and so forth do actually subsist," *Principles* § 47 tells us that "the matter philosophers contend for is an incomprehensible somewhat, which hath none of those particular qualities

whereby the bodies falling under our senses are distinguished one from another." In these two passages taken together, Berkeley is not nailing down an inconsistency in Locke; he is indulging in an inconsistency which arises from his misunderstanding of Locke's problems.

The literature does not yield the same rich harvest of thorough, glad commission of this conflation as it does for the one discussed in the first part of my paper. Most commentators merely take Berkeley's word for it that the veil-of-perception doctrine is integrally connected with the thesis about primary and secondary qualities, and lurch somehow across the gap where the connection is supposed to be. In somewhat the same way, they accept but do not intelligently argue for Berkeley's demonstrably false claim that Locke's theory of abstract ideas is connected with the three-headed monster which Berkeley calls Locke's doctrine of material substance.

It is a disaster that the British empiricists, and Berkeley in particular, have received such cursory attention in recent years. There is so much in them that is boldly and energetically wrong that they can be enormously instructive as object lessons, if in no other way, to anyone who will attend *closely* to their thought. The wretched course of English epistemology over the past half-century illustrates, depressingly, the dictum that those who do not study history are in danger of reliving it.

University of Cambridge

II. THE HIGH ROAD TO PYRRHONISM*

RICHARD H. POPKIN

THE advent of René Descartes' attempted resolution of the "sceptical crisis" of the Renaissance and Reformation posed a crisis of sorts for the "nouveaux Pyrrhoniens," the new sceptics who constituted the avant-garde of the French intellectual world, and the heirs of Montaigne. From Montaigne and his cousin, Francisco Sanchez, through Pierre Charron, Bishop Jean-Pierre Camus, Gabriel Naudé, François de La Mothe Le Vayer, and others, the epistemological arguments of ancient Greek Pyrrhonism had been revitalized as means of attacking Renaissance Platonism, Scholasticism, Calvinism, astrology, alchemy, and a host of other views, ancient and modern. Descartes, raised in this context, outdoubted his contemporaries in order to find a truth so certain that all of the most extravagant suppositions of the sceptics could not shake it. Then, he went on to erect a new type of dogmatic philosophy that was to affect all subsequent efforts to make our intellectual universe coherent and consistent.¹

In the face of this new philosophy, the sceptics had to redirect and re-examine their arguments. Descartes had, to an extent, stolen their thunder, by taking over their arsenal, then introducing a new super-sceptical weapon, the demon hypothesis, to annihilate the same opponents. But, then, having accomplished all of this, Descartes had dared to claim that he had found certainty through scepticism, and was ready to give indubitable answers to all of the major questions of philosophy, and to erect a new edifice of human certitude. One group of anti-Cartesians, late Scholastics such as Father Bourdin, Pierre Petit, and Gisbert Voetius, insisted that Descartes was just a sceptic, and that, all of his claims to the contrary notwithstanding,

he had only succeeded in further undermining the bases of human knowledge. Voetius and Schook at Utrecht saw the Father of Modern Philosophy as at least a partial and secret sceptic.²

The avowed sceptics, on the other hand, were more impressed by the positive claims from the *cogito* onwards than by the method of doubt. Some of them recognized immediately that a new challenge had been made, and that it was their duty to their cause to show the preposterousness of the new dogmatism of René Descartes. The development of scepticism after Descartes—what I have labeled with Bayle's phrase, "The High Road to Pyrrhonism"³—will be the subject of this paper, leading from the problem of redirecting the classical Greek arguments to the complete unraveling of the warp and woof of the intellectual world by Pierre Bayle. It is during this 50 year period, I believe, that the basic structure of the original "new" philosophy was dissected, analyzed, and destroyed, leaving wounds of such magnitude that philosophy has never really recovered from them. The marvelous optimism with which Descartes launched the modern intellectual world was to be gradually so completely eroded during the second half of the 17th century that philosophy could carry on thereafter only by changing its character, and denying its previous role.

It is interesting to note that some of the old guard sceptics, like Gabriel Naudé, and La Mothe Le Vayer, hardly bothered taking note of the Cartesian revolution. They disliked Descartes as a person because of his dogmatic self-assurance. They imbibed an anti-Cartesian attitude from their friends, Hobbes and Gassendi. But in their writings, their scepticism was still just an insipid reflection of their idols, Montaigne and Charron. La Mothe

* This paper was originally presented as a lecture at the annual meeting of the Society for French Historical Studies at Harvard in April 1963.

¹ For details on the development of scepticism during the Renaissance and early 17th century, see R. H. Popkin, *The History of Scepticism from Erasmus to Descartes* (Assen, The Netherlands, 1960); "Scepticism et Contre-Réforme en France," *Recherches et Débats du Centre Catholique des intellectuels français*, Cahier No. 40 (*Essais sur Teilhard de Chardin*), October 1962, pp. 151-184; and "Scepticism and the Counter-Reformation in France," *Archiv für Reformationsgeschichte*, vol. 51 (1960), pp. 58-87.

² Cf. Popkin, *History of Scepticism*, chap. X, pp. 196-202.

³ Pierre Bayle uses this phrase with regard to the Catholic "way of authority" for establishing religious truth in his article NICOLLE (Pierre), Rem. C, of the *Dictionnaire Historique et Critique*.

Le Vayer lived more than 20 years after Descartes' death, but kept turning out essentially the same scepticism, with the same Montaignian arguments, against the same old opponents.

But, at the same time that the more literary sceptics had not noticed that the intellectual world had changed its character, the more scientifically minded ones did. Pierre Gassendi, who had started his career by marshaling all of the sceptical ammunition he found in Sextus Empiricus, Montaigne, Sanchez, and Charron, against Aristotelianism and Renaissance Platonism proclaiming that no science is possible, least of all in Aristotle's sense, leaped into the fray against Descartes almost immediately. In his objections against Descartes' *Meditations*, then expanded into the enormous *Disquisitio metaphysica*, Gassendi brushed aside Descartes' sceptical side, and set to work on his dogmatic one. Using the same classical sceptical arguments, Gassendi began a new phase in the history of scepticism, by redirecting and revising the arguments so that they now aimed directly at the Cartesian theory of knowledge. The *cognito* and the new criterion of truth, Descartes' doctrine of clear and distinct ideas, were now the targets. Gassendi challenged whether *cogito ergo sum* was a serious proof of anything, whether the truth involved was more than an obvious triviality, and whether anything followed from it. He tried to show that the classical sceptical problems about knowledge could be raised against the Cartesian theory, by asking for a criterion by which we could tell when some "truth" was really clear and distinct, and not just that it appeared to be clear and distinct. Gassendi saw that the very same issues that were raised by the ancient Pyrrhonists to baffle their Stoic and Epicurean opponents could be introduced in a new form against Descartes.⁴ And Descartes, in spite of his low opinion of Gassendi, appreciated the force of this attack to the extent of labeling one of Gassendi's gambits "the objection of objections." Gassendi had raised the problem of whether we could ever tell that the world of clear and distinct, and hence certain, ideas is just a creation of our mind, having no relation to reality. If we could not tell, then all

of our vaunted "knowledge" might turn out to be only about how things seem to us, and not about how they *really* are. Descartes' answer to the "objection of objections" is only to point out the catastrophic results that would follow if one took this sceptical problem seriously. If it were the case that all that we could ever know were only thoughts in our own minds, "it follows that there is nothing that we can in any way understand, conceive or imagine that should be considered as true. That is to say, that we have to completely shut the door on reason, and content ourselves with being monkeys, or parrots, and not men any longer."⁵

This kind of sceptical attack against Descartes was carried on to a lesser degree by Samuel Sorbière, Gassendi's disciple, and to a higher degree by Bishop Pierre-Daniel Huet and Simon Foucher. Sorbière lived in Holland during the latter part of Descartes' stay there, and had been raised on Sextus Empiricus, while accepting the nonmetaphysical, and quasi-empirical sides of his friends, Gassendi and Hobbes. As a result, he saw Descartes as a man who had gone too far in both directions, into useless doubting, and preposterous affirming. For Sorbière, Descartes had missed the whole point in scepticism, which was to clear away the quest for knowledge of reality, so that one could concentrate on serious matters, the careful, cautious scientific study of experience. The doubts should not be raised just once, and employed as a universal purge to eliminate all opinions and beliefs, but should be raised constructively as a daily dosage, to be taken in small quantities (Sorbière, like Sextus, was a physician by profession) for the health of the mind, as a quiet and benign remedy which protects one from poorly based opinions. But, doubts should not be employed, à la Descartes, as a poison that destroys even the first principle of our reasoning.⁶

In a still unpublished letter on the death of Descartes, Sorbière gives his careful evaluation of the accomplishments of the great man and concludes that he was a first-class mathematician, but that his scientific achievements were not so good. He [Descartes] impressed the "demi-curieux" by setting forth formulas and demonstrations. But

⁴ See Pierre Gassendi, *Disquisitio Metaphysica*, ed. by Bernard Rochot, with French translation (Paris, 1962), esp. the sections on Descartes' 2nd and 3rd Meditations; Gassendi, *The Fifth Set of Objections*, in René Descartes, *Philosophical Works*, Haldane-Ross ed. (New York, 1955), vol. 2, esp. pp. 151-152 (Adam-Tannery ed., vol. VII, pp. 278-299).

On Gassendi, see Popkin, *History of Scepticism*, chap. V and chap. VII and references given there.

⁵ Descartes, *Lettre de Monsieur Des-Cartes à Monsieur C.L.R.* (Clerelier), in Adam-Tannery ed. of *Oeuvres*, vol. IXA, pp. 211-212, translated in Haldane-Ross, vol. 2, p. 131.

⁶ Samuel Sorbière, *Lettres et Discours de M. de Sorbière sur diverses matieres curieuses* (Paris, 1660), lettres LXXXVII and LXXXVIII, esp. pp. 690-691 in the latter.

since the study of nature isn't like mathematics, and deals in facts and not abstractions, he has not actually shown anything about reality. Sorbière insisted that his friends, Gassendi and Hobbes, had accomplished much more because they kept, most of the time, to natural science as the empirical study of the world.⁷ Sorbière, in this respect, was so empirical that he managed to miss the point of the best contributions of his heroes, as well as of the new experimental work of the Royal Society of England. Sorbière chided Gassendi for offering hypotheses about unseen events. When Gassendi, in keeping with his atomic theory (which he proposed *solely* as an hypothetical model to connect experienced events), proposed a theory about the circulation of the blood, Sorbière wrote that he preferred to restrict himself to the world of appearance alone.⁸ On the other hand, Sorbière was the translator of Hobbes' *De Cive* (a task he undertook when Father Mersenne told him that if he examined this work it would cure his scepticism). *De Cive* he found completely to his taste since he saw it only as a science of visible men, and visible societies.⁹ He failed to see the powerful theoretical tools that both Gassendi and Hobbes were developing which were to inspire so much of the underlying theory of modern physics and the behavioral sciences, and he admired his heroes instead for their empirical and mathematical work, which was their weakest point. And, to finish off his story, after Sorbière went to England in 1663, becoming one of the first two foreigners elected to the Royal Society, he returned to France and wrote a wonderful exposé of the whole trip, offering his comments on how bad the food was, how bad the hotels were, and how bad everybody's Latin was. After attending the infant Royal Society, he decided that

Hobbes, who was sulking in exclusion from that august body, was right. They were just tinkerers, and inferior mathematicians, but no match for England's *real* scientist—Thomas Hobbes. Bishop Sprat, in angry reply for the Royal Society, suggested that if Sorbière was really the sceptic he claimed to be, he should not be so dogmatical about the lack of merit of *la cuisine anglaise*, but should remain in suspense of judgment on the issue.¹⁰

Sorbière saw the Cartesian revolution as a totally misguided venture, both in its negative and its positive aspects. He tried to ignore it in favor of what was to be the wave of the future—empirical science. His own misunderstanding of the nature of empirical research led him to miss the real contribution of his friends, and to mistake the locus of the fruitful experimenters. But, he saw one major aspect of the sceptical development vis-à-vis Descartes—that it would provide a substitute for what Descartes was trying to achieve—a science shorn of any metaphysical foundations, to be assessed only on its predictive and practical results.

Two other sceptical figures, Bishop Huet and Simon Foucher, brought about the next stage of this history—the final destruction of the Cartesian system through attacking it with the use of all of the epistemological weapons of classical scepticism. Foucher (1644–96), canon of Dijon, was originally in close contact with the Cartesian circle in Paris. He was chosen to give the funeral oration for Descartes when the great man was finally buried in the 1660's on holy French soil. Foucher, around 1667, started raising embarrassing sceptical problems for the Cartesians about how we could possibly gain knowledge of the external world

⁷ Letter of Sorbière to Saumaise, La Haye, 10 Mars 1650, Bibliothèque Nationale Ms. Fr. 3930, fol. 262.

"Pour nostre defunct [Descartes], je vous advoue que ce peu que je scay de Mathematiques me le fait reconnoistre l'un des premiers hommes du monde, en Algebre & en Geometrie; & je l'ay ouy renommer pour tel à Roberval, Bonnel, Hobbes & Fermat, qui sont des plus grands maistres. Mais quand il s'est meslé de Physique il n'a pas mérité la même louange; quoy que les demi-curieux la luy ayent donnée. Ils se sont laissés surprendre à quelques lignes; & ont jugé, à cause qu'ils voyoient des figures, & que les Mathematiques qui en usent, sont certaines, que tout ce qu'il disoit estoit autant de demonstrations. Mais en cela ils se sont abusés, selon mon advis, aussi bien que ce subtil Philosophes qui n'a pas assez considéré, qu'il n'en est pas de la science naturelle de même que des Mathematiques dans lesquelles il est permis de supposer des choses fausses, parce que leurs suppositions rapportent aucune inconvenient & qu'elles sont au contraire fort commodes aux demonstrations qui en sont tirées; Là où dans la Physique on procede un peu plus grossierement; on ne raisonne pas sur des abstractions mais sur les experiences des sens, ou sur des choses un peu plus grossieres ou plus palpables que ces estres auxquels il faut qu'un grand effort de l'imagination donne quelque realité."

⁸ Sorbière, *Discours sceptique sur le passage du chyle, & le mouvement du coeur* (Leyden, 1648), pp. 153–154. Cf. Gaston Sortais, *La Philosophie moderne depuis Bacon jusqu'à Leibniz* (Paris, 1922), tome II, p. 194.

⁹ Cf. Sortais, *op. cit.*, vol. II, pp. 214–216.

¹⁰ Cf. Vincent Guilloton, "Autour de la *Relation* du Voyage de Samuel Sorbière en Angleterre 1663–64," *Smith College Studies in Modern Languages*, vol. 11 (1930), pp. 1–29; and Thomas Sprat, *Observations on Monsieur de Sorbière's Voyage into England* (London, 1665), pp. 275–276.

through clear and distinct Cartesian ideas. In the early 1670's Foucher came to know Leibniz, and through discussing his sceptical views, he not only developed some ingenious sceptical ploys, but he also led Leibniz to spend a good deal of time and energy working out a counter-theory (some of the earliest and most important statements of Leibniz' philosophy originally appeared as answers to Foucher). When Father Malebranche appeared on the scene in 1674, with *De la Recherche de la Vérité*, Foucher rushed into print with his sceptical attack, partly because he had been writing a book with approximately the same title, and partly because he misread Malebranche, and thought he was just another "party-line" Cartesian. From then on Foucher spewed out volumes with titles like *Critique de la Recherche de la Vérité*, *Dissertations sur la Recherche de la Vérité*, etc., in which he kept presenting his sceptical attack against Cartesianism, and his revival of the classical scepticism of the New Academy (set forth, always, as true Augustinianism. Foucher contended that Augustine's *Contra Academicos* was really *pro Academicos*). Malebranche was unimpressed and regarded Foucher as a fool, since he couldn't distinguish Malebranche's innovations from standard Cartesianism (while an orthodox Cartesian like Arnauld was to point out that these innovations led to a most dangerous Pyrrhonism). Others also seem to have regarded Foucher as a fool, and even Leibniz, after a while, was to have his doubts about his friend's intellectual abilities. Foucher's major contribution was, I believe, based on a misunderstanding of both Descartes and Malebranche, but, for better or worse, it was to become a major ingredient in subsequent anti-rationalistic philosophy.¹¹

Foucher's contribution centers on challenging the central Cartesian claim that some of our ideas, those of the mathematical or primary qualities, can give us knowledge of the external world, while those of sensory or secondary qualities (tastes, smells, colors, sounds, etc.), cannot. Descartes and all others (Galileo, Hobbes, Spinoza, Malebranche, Locke, etc.), of the "new philosophers" had dis-

tinguished primary and secondary qualities, had offered standard sceptical reasons for contending that secondary qualities were entirely subjective, that is, in the mind, whereas primary qualities were objective, that is, were qualities of the real external world. (And hence that a mathematical physics of the real world was possible.) Bayle had noted that modern philosophy began with the ancient Pyrrhonist Sextus Empiricus, who provided the reasons for doubting the reality of secondary qualities.¹² Hume later said that the fundamental principle of all of what he called the "modern philosophy" "is the opinion concerning colors, scounds, tastes, smells, heat and cold: which it asserts to be nothing but impressions in the mind, deriv'd from the operation of external objects, and without any resemblance to the qualities of the objects. . . . This principle being once admitted, all other doctrines of that philosophy seem to follow by an easy consequence. For upon the removal of sounds, colors, heat, cold and other sensible qualities from the rank of continu'd independent existences, we are reduc'd merely to what are called primary qualities, as the only *real* ones, of which we have any adequate notion."¹³

Foucher challenged this fundamental principle in two ways, one by questioning if *any* ideas could give us knowledge of external objects, and the other by using the *same* sceptical arguments that established the subjectivity of secondary qualities to establish the subjectivity of primary ones. The first way (which Huet was to develop in expert fashion) involves pointing out that ideas are presumably not like external objects and then asking how can ideas tell us about non-ideas? The stock Cartesian line is that "ideas represent objects." This, Foucher and then Huet insisted is nonsense. Ideas are not like objects; they don't resemble objects, so how can they represent objects? Nothing can be like an idea but an idea! When Foucher and Huet were told *ad nauseum* by Cartesians that "representer" should be taken to mean "to make known," and not "to look like," they replied that this doesn't help, for how can an idea lead to or explain anything but an idea? (All of

¹¹ On Foucher, see F. Rabbe, *Étude philosophique sur l'abbé Simon Foucher* (Paris, 1867); Henri Gouhier, "La première polemique de Malebranche," *Revue d'Histoire de la Philosophie*, vol. 1 (1927), pp. 23-48 and 168-191; R. H. Popkin "L'abbé Foucher et le problème des qualités premières," *XVII^e Siècle* No. 33 (1957), pp. 633-647; and Richard A. Watson, "The Breakdown of Cartesian Metaphysics," *Journal of the History of Philosophy*, and his article on Foucher in *The Encyclopedia of Philosophy*, forthcoming. Dr. Watson, a former student of mine, has just completed a book length study on Foucher, and finds his views more cogent than did previous commentators from Leibniz and Malebranche down to Popkin. *Vide* Watson, "The Breakdown of Cartesian Metaphysics," *Journal of the History of Philosophy*, vol. 1 (1963), pp. 177-197.

¹² Bayle, *Dictionnaire*, art. *Pyrrhon*, Rem. B.

¹³ David Hume, *A Treatise of Human Nature*, Selby-Bigge ed. (Oxford, 1951), Bk. I, Pt. IV, § iv, pp. 226-227.

335076

this, for obvious historical reasons, should bring some of Bishop Berkeley's arguments to mind.)¹⁴

Foucher's other main argument was that the same sceptical evidence that makes all the Cartesians say that the secondary qualities are "in the mind" should also make them say the same of the primary qualities. All of the qualities we are aware of, he argued, are known by direct perception, that is, by sensations. Sensations are only modifications of our souls, "from which it follows that if one admits that we know about extension and shapes by the senses as well as about light and colors, it will be necessary to conclude that this extension and these shapes are no less in us than that light and those colors."¹⁵ If it is admitted that our senses deceive us regarding the objective existence of secondary qualities, then the Cartesians have to show that the same is not the case with regard to the primary qualities. When the Cartesians claim that God would deceive us if there were no extended world outside of us, the same would be true with regard to the non-existence of a colored, or an odoriferous, or a noisy world.¹⁶

Once it is admitted that the secondary qualities are "in us," how can one avoid reaching the identical conclusion with regard to primary qualities? The extended object that we perceive is also colored. "You recognize that these colors are in us! But where is the shape of these colors, the extension of these colors if not in the place where the colors are?"¹⁷ The very same sceptical arguments about the variations of sense experience that have made all new philosophers say that secondary qualities are in us, and not in the external world, show that the same applies to the perceived primary qualities.

With these arguments, Foucher thought that he had shown that neither Descartes nor Malebranche

had a theory of knowledge that could lead to any information about an external real world, and hence that Cartesianism could not provide a basis for any certainty about the objective physical universe. The Cartesians could point out (and did) that they never said that primary qualities were known by sense experience (to which Foucher replied that no other means of knowing them seemed likely, nor would any other help, since the knowledge would consist of ideas, which would still be modifications of the mind, with the problems this raised.) Malebranche could point out (and did) that he didn't hold to the Cartesian theory of knowledge on this point (to which Foucher replied that Malebranche only seemed to achieve making any knowledge of anything incomprehensible and miraculous.) But, no matter what Foucher understood or misunderstood of his opponents' view, his arguments were to reverberate through the subsequent history of philosophy as major achievements of Berkeley and Hume. The arguments passed into the main stream of ideas through Bayle, who was one of the few readers of Foucher's early works. In the *Dictionary*, article "Pyrrho," remark B, and in article "Zeno," remarks G and H, Bayle cites Foucher's arguments and embellishes them, as the means by which scepticism destroys the "new philosophy." Both Berkeley and Hume studied these Baylean discussions and used them as their own dialectical tools against all the previous non-empirical theories of knowledge.¹⁸

Foucher provided some crucial refurbished sceptical arguments to meet the developing new dogmatism—Cartesianism. Many more were provided by his and Leibniz' friend, Bishop Pierre-Daniel Huet, 1630–1721. Huet started off as a classical scholar, and apparently a Cartesian of sorts. He was in Stockholm shortly after Descartes'

¹⁴ Simon Foucher, *Critique de la Recherche de la vérité* (Paris, 1675), pp. 50–52; *Réponse pour la Critique à la préface du second volume de la Recherche de la vérité* (Paris, 1676), pp. 42–43, 54, and 112; *Dissertation sur la Recherche de la vérité, contenant l'apologie des Académiciens* (Paris, 1687), pp. 86–87; and *Nouvelle Dissertation sur la Recherche de la vérité* (Paris, 1679), pp. 30–40. Pierre-Daniel Huet, *Censura Philosophiae Cartesianae* (Paris, 1689), Chap. II; and *Traité philosophique de la faiblesse de l'esprit humain* (London, 1741), livre I, chap. iii. See also the two works by Huet's friend, Jean Du Hamel, *Reflexions critiques sur le système Cartésien de la philosophie de M. Régis* (Paris, 1692), pp. 27–35, and *Lettre de Monsieur Du Hamel, Ancien Professeur de Philosophie de l'Université de Paris, pour servir de Réplique à Monsieur Régis* (Paris, 1699), p. 16.

For the Cartesian side of the story, one of the best statements appears in Pierre-Sylvan Régis, *Réponse aux Reflexions critiques de M. Du Hamel* (Paris, 1692), pp. 8–20, where Régis developed the thesis that ideas make things known, but that ideas do not resemble things. See also Régis' *L'Usage de la raison et de la foy* (Paris, 1704), pp. 23 ff.

¹⁵ Foucher, *Critique de la Recherche de la vérité*, pp. 61–79. (The pagination jumps from 66 to 77. The quotation is on p. 79.)

¹⁶ Foucher, *Réponse pour la Critique*, p. 113, and *Nouvelle Dissertation*, pp. 42–43.

¹⁷ Foucher, *Nouvelle Dissertation*, p. 46.

¹⁸ On the controversies over Foucher's views, and on the historical impact of his arguments, see Popkin, "L'abbé Foucher et le problème des qualités premières," esp. pp. 638–647; "Berkeley and Pyrrhonism," *Review of Metaphysics*, vol. 5 (1951–52), esp. pp. 227–231; and Watson, "The Breakdown of Cartesian Metaphysics."

death there, and came in contact with some of the sceptical anti-Cartesians. Next he was a minor figure in Naudé's *libertin* circle in Paris, and a constant friend and ally of the growing group of Jesuit anti-Cartesians. Finally Huet was appointed to half of the successorship of La Mothe Le Vayer, the heir of Montaigne, twice removed, and teacher of the Dauphin. Huet served with Bossuet as a preceptor in the Court until the Dauphin no longer needed or could profit from their instruction.¹⁹ During this period, Huet did the enormous amount of work that convinced Leibniz and others that he was the most learned man alive. Two of his most sensational works, his *Censura philosophiae cartesianae* and *Demonstratio Evangelica*, are from this period. At the age of 50, Huet was ordained and rewarded for his services at Court with appointment as Bishop of Soissons, a place he disliked, which he traded for the Bishopric of Avranches. After a few years on the job, where his parishioners complained that he was always too busy studying to take care of the diocese, he retired and lived out most of his remaining long life in a Jesuit house in Paris.²⁰ (It is of note that during both his political and ecclesiastical career, he seemed to be the only major figure not actively involved in either the Jansenist or the Calvinist persecutions. He himself was from a Protestant family that had been converted, and many of his best friends were Norman Calvinists. During the worst of the persecutions, he gave aid and comfort to friends and relatives of friends. Besides his astounding latitudinarianism in his theological works, his major ecclesiastical contribution seems to have been to try, with Leibniz, to reunite all the churches by eliminating all the points of disagreement from the essence of Christianity.) Huet's most famous work, the sceptical *Traité philosophique de la faiblesse de*

l'esprit humain, written around 1690-91, was not printed until after his death, and caused a great sensation then, when it was revealed, in a period when this no longer seemed plausible or proper, that a leading French Bishop was a complete Pyrrhonian sceptic.²¹

Huet devoted a great deal of his intellectual energy to training the traditional sceptical epistemological arguments on two basic Cartesian points, the fundamental truth of the *cogito* and the criterion of clear and distinct ideas. In the *Censura* and in a still unpublished defense of it, Huet analyzed every element of the *cogito* to show that the alleged indubitable truth, "I think, therefore I am" was actually a somewhat dubious claim. Besides developing the arguments that Gassendi and Hobbes had earlier started about the legitimacy of the inference from "I think" to "I am," Huet pointed out that the entire proposition asserted in the *cogito* represented a thought sequence. By the time one comes to the end of the proposition the beginning is now a remembered mental event. Since memory is fallible, it might not be the case that one thought. When one is aware that one thinks, the conclusion is still in future, and one does not yet realize that one is. Hence, Huet contended, the *cogito* when properly stated reads "I may have thought, therefore perhaps I may be." Huet's dissection of the *cogito* was joined with an equally elaborate examination of the Cartesian criterion of truth. In his dispute with the last major Cartesian, Pierre-Sylvan Régis, Huet employed all the weapons that his erudite study of the earlier sceptics had turned up, to dissolve the Cartesian theory of knowledge.²² In concert with his fellow sceptics who devoted themselves to attacking Cartesianism, and his allies among the Jesuit anti-Cartesians, like René Rapin, Gabriel

¹⁹ Actually, in poring over the fantastic amount of unpublished Huet materials that exist all over Europe (especially the collections in the Bibliothèque Nationale of Paris, in the Biblioteca Laurenziana of Florence, and the Bibliothèque de la Ville of Caen), I have been impressed by the fact that there are a great many bundles of the Dauphin's homework, and a great many indications that wherever the Court was, Huet wasn't. If they were at Saint-Germain, he was in Paris, and so on. He was constantly too sick or too busy to give the Dauphin the requisite instruction. See, for instance, the letters from the Duke of Montausier to Huet in the Carteggio Huet, Ashburnham Collection, 1866, Biblioteca Laurenziana, Florence, Cassetta VIII, items 1429-1600; and Bibliothèque Nationale, Ms. Nouv. Acq. 6202, fols. 47 ff. on the "devoirs de Dauphin."

²⁰ The most complete study of Huet's life and career is that by the Abbé Léon Tolmer, *Pierre-Daniel Huet (1630-1721), Humaniste-physicien* (Bayeux, 1949). This work unfortunately does not incorporate the wealth of material in the Ashburnham papers in Florence, and so is incomplete on certain matters.

²¹ It was first published in 1723, and soon appeared in English, Latin, and German versions as well. See Tolmer, *op. cit.*, pp. 549-553 for the reaction when it was published. The authenticity of the work was challenged, and leading scholars had to attest that they had seen it in manuscript in Huet's own hand.

²² Cf. Pierre-Daniel Huet, *Censura Philosophiae Cartesianae*, Cap. i-ii; and *Censure de la réponse faite par M. Régis au livre intitulé Censura philosophiae Cartesianae, par Theocrite de la Roche, Seigneur de Pluvigny*, an unpublished manuscript, Bibliothèque Nationale, Paris, Ms. Fr. 14703, no. 3.

Daniel, and Louis de La Ville,²³ Huet advocated a probabilistic nonmetaphysical study of the world as the way to proceed. He saw the experimental philosophy of the Royal Society as the proper attitude for a sceptic,²⁴ instead of the arrogant, illiterate, plagiarized, dogmatic view of the late René Descartes. (Huet and his crony Leibniz were delighted every time they could uncover a text from which they believed Descartes pilfered, or could show that Descartes did not read Greek, or was unaware of some scholarly news. In his satire about what happened to Descartes in Lapland after his supposed death in Stockholm, Huet heaped endless ridicule on the thought and character of the Father of Modern Philosophy.)²⁵

In his positive views, Huet, like Foucher, Gassendi, and Sorbière advocated an empirical probabilistic approach in the sciences and an acceptance of religion on faith alone. His most influential work, the *Demonstratio Evangelica*, grew out of his conversations with Rabbi Menasseh ben-Israel in Amsterdam. In it, he purports to argue first, that the evidence for religion is as good as that for mathematics, since the latter discipline is uncertain, and at best only probable. Then, Huet attempted to develop a grandiose induction to show the evidence for Christianity, an induction based on finding common elements in all known religions, and then claiming that Christianity was both the same as every other religion and was the right formulation of the common elements. From his vast erudition, he found a Moses figure, a Jesus figure, etc., in every known religion, ancient and modern, and then he insisted the Christians alone had all the names correctly spelled and the story in the right order. The enormous collection of evidence he adduced was to provide all the anti-

Christians of the next century with endless data to show that Christianity was just the same as all the other (false) religions. So that instead of inducing the characteristics of the true religion from comparative studies, the Enlightenment arguers denied the truth of all historical religions on the basis of Huet's data. Perhaps Huet came closer to stating his real theological view when he defended someone who had maintained the theses that it is not evident that there is a true religion, and that it is not evident that the Christian religion is the most probable one. Huet reported that he was in accord with these theses, though they might, by inadvertency, give the impression that the holder of them was not a true and believing Christian. If Christianity is based on faith alone, then, Huet announced, there should be no rational evidence of the existence of God, and no rational probability that Christianity is true.²⁶ And, in his notes on Pascal, Huet could only find the fideism of the *Pensées* too rationalistic, because of the wager argument.²⁷

The sceptics Gassendi, Sorbière, Foucher, and Huet all directed their fire primarily against the new dogmatic menace, Cartesianism, and principally against its theory of knowledge. They, along with Pascal, and the Jesuit anti-Cartésians, revealed that the classical sceptical artillery could reduce the new Cartesian edifice to a shambles, and various crucial citadels could be destroyed by readjusting and reworking the traditional sceptical gambits. In the light of these objections, the long controversy between Arnauld and Malebranche showed that neither the original nor the revised form of Cartesianism could escape the sceptical difficulties, and that either form became either a "dangerous Pyrrhonism" or a "ridiculous Pyrr-

²³ Cf. Gabriel Daniel, *Voyage du monde de Descartes* (Paris, 1690); Louis de La Ville, *Sensiments de M. Descartes, touchant l'essence & les propriétés du corps*, . . . (Paris, 1680); and René Rapin, *Reflexions sur la philosophie ancienne et moderne* (Paris, 1676). (Rapin was a close friend of Huet's.)

On all of these figures, and other Jesuit anti-Cartésians, see Leonora Cohen Rosenfield, "Peripatetic Adversaries of Cartesianism in 17th Century France," *Review of Religion*, 1957, pp. 14-40; and Gaston Sortais, "Le Cartésianisme chez les jésuites français," *Archives de Philosophie*, vol. VI, Cahier III (Paris, 1929).

²⁴ This view is developed in Huet's *Traité*, especially in Livres II and III. Livre II, chap. X praises the work of the Royal Society. Rapin presented a similar view in his *Reflexions sur la philosophie ancienne et moderne*.

²⁵ See Huet's *Censura*, and his many marginal notes in his copies of Descartes' works in the Bibliothèque Nationale. The copy of the *Censura*, BN Rés. R. 1963, has lots of marginal attacks on Descartes on these grounds, and on p. 218v gives the text of a letter of Leibniz of Dec. 19, 1689, attacking Descartes. Huet's *Nouveaux Mémoires pour servir à l'histoire du Cartésianisme* (Paris, 1692), also contains many jibes and personal attacks on Descartes.

²⁶ Cf. Huet, *Demonstratio Evangelica*, 3rd ed. (Paris, 1690); and Biblioteca Laurenziana Ashburnham Ms. 1866 (Carteggio Huet), Cassetta XIII, no. 2435 (Lettre de Huet à le P. l'Honoré, Avranches, 27 Mars 1693), and Cassetta XV, no. 2899, "Lettre d'un Docteur de Paris à un docteur de Caen touchant l'évidence de la Religion Chrétienne." This latter document analyzing the thesis (not by Huet) states that these theses were sustained on Jan. 30, 1693, at the Jesuit University of Caen.

²⁷ These notes are in Huet's copy of the *Pensées*, Bibliothèque Nationale Rés. D.21375. There is a long analysis and critique of the wager argument on the back flyleaf, dated 23 Février '70. (The *privilege* in this edition is dated Jan. 2, 1670, so Huet's comments must have been written right after he obtained the work.)

honism."²⁸ In the half century after the publication of the *Meditations* the sceptics had been able to counter-attack and question each crucial ingredient of the Cartesian theory of knowledge, and to raise apparently insoluble doubts about the value of the entire Cartesian undertaking.

All of the sceptics discussed so far were Catholics (except for Sorbière, who remedied this defect by conversion) and stated a common view that the purpose or result of their scepticism was a "defense" of the Faith. By destroying the rational aspirations of mankind through undermining various philosophies and theologies, man's only recourse would be to faith. They usually took Calvinism to be a new form of rational dogmatism, offering reasons for changing one's religion. The exhibiting of the impossibility of man's finding the truth by reason was, they claimed, the road to faith, and a "justification" for remaining in the Catholic camp (since no reasons existed for leaving it). Some of these sceptics saw Descartes and Calvin as the twin enemies—Descartes the new dogmatic rational philosopher, Calvin the new dogmatic, rational theologian. Elsewhere I have dealt with the sceptical Counter-Reformation. Let it suffice to say here that the sceptics, especially Huet and Foucher, attacked Cartesianism as a new menace to true religion and claimed that by destroying it they were preserving the true faith. Huet joined his contention to the anti-Cartesian campaign of the Jesuits, and they mingled their critical anti-Cartesian views into a compound view—that Cartesianism was indefensible because it could be challenged by the traditional sceptical arguments, that Cartesianism was irreligious, Calvinistic, etc., etc. It should be replaced, they all agreed, sceptics and Jesuit anti-Cartesian in unison, not by a different older or newer rational philosophy and, theology—Aristotelian, Scholastic or neo-Scholastic—but by *no* metaphysical view of the nature of reality. An empirical study of nature and a fideistic acceptance of the faith should go hand in hand. Gassendi was to be preferred as a guide to either Aristotle or Descartes, with regard to the

natural world, and Augustine and Tertullian to any rational theologians.²⁹

Huet and Foucher, and their Jesuit allies, destroyed Cartesianism at its epistemological heart. Ostensibly, this was done in the service of the Faith. The critiques they set forth are sufficient to undermine the core of the new way of ideas introduced by the new philosophy. Though they may all have been sincere in the Counter-Reformation zeal (and I think that they all were), the effect of their anti-racial endeavors was to unleash a form of "rational" irreligion in the next century. Huet's valiant efforts to defend the unreasonableness of religion and to expose the hopelessness of the way of ideas, and Foucher's valiant efforts to show that Cartesianism could lead to no knowledge of reality, were to be taken up by Berkeley and Hume and the French Enlightenment as reasons for rejecting nonempirical philosophies and the Judeo-Christian tradition. Foucher's ploy was to overturn modern philosophy by destroying the ontological distinction between primary qualities (of reality) and secondary qualities (existing only in the mind). Huet was to break down any Cartesian means of gaining knowledge, and to amass a body of data about the relativity of the Judeo-Christian tradition to a localized part of human history and culture.

These efforts, combined with the more generalized sceptical attack by Pierre Bayle, were to undermine all rational endeavors altogether. Bayle, 1647–1706, was to put together and add to the critique of Cartesianism, to carry the sceptical onslaught to all types of theories—of knowledge, metaphysics, theology, and science, and to unravel the warp and woof of man's intellectual world. The 17th-century sceptics before Bayle were concerned primarily with epistemology, and, as I have tried to indicate above, with directing the traditional Greek sceptical arguments about the possibility of knowledge against the new dogmatism of Descartes. Basically, the scepticism that runs from Gassendi to Huet is an epistemological scepticism aimed principally at showing that the

²⁸ See, for example, Antoine Arnauld, *Defense de Mr. Arnauld Docteur de Sorbonne contre la Réponse au livre des vraies et des fausses idées* (Cologne, 1684), pp. 577–578; "8^e Lettre de Arnauld au R. P. Malebranche," in *Oeuvres de Messire Antoine Arnauld*, Tome XXXIX (Lausanne, 1781), p. 133; and Nicolas Malebranche, "Réponse du Père Malebranche, Prestre de l'Oratoire, à la troisième lettre de M. Arnauld, Docteur de Sorbonne . . .," in *Recueil de toutes les Réponses du P. Malebranche à M. Arnauld*, Tome IV (Paris, 1709), pp. 51–52.

²⁹ The Jesuits Gabriel Daniel and René Rapin expressed preference for Gassendi over most of the other moderns. The tenor of Rapin's *Reflexions* is that Aristotle should not be blamed so much since *no one* can gain true and certain knowledge of nature. Descartes and other modern metaphysicians are culpable since they are leading people away from constructive empirical science and sound (Augustinian) theology.

Cartesian claim to have found a new basis for, and way to, knowledge is untrue, or at least, highly dubious. Bayle not only took over this aspect of 17th-century scepticism, but he made it part of a massive, all-encompassing attack on all attempts to comprehend and explain man's world. Bayle's aim was to undo the very effort to find rationality in the universe, partly by using the methods of previous sceptics, partly by unleashing the attacks of Sextus Empiricus on metaphysics, mathematics, ethics, etc., as well as on epistemology, against all theories, new or old, and partly by incorporating other critical techniques into the sceptic's arsenal. Bayle's efforts were to provide the high road to complete Pyrrhonism, undermining completely the brave new world of 17th-century metaphysics, and leaving man, in spite of his claims to the contrary, afloat in "a sink of uncertainty and error."

To appreciate Bayle's accomplishment, I think one has to see him not as the culmination of his predecessors, but as a unique figure who came from a quite different background from that of the Montaigne tradition. Bayle was a Protestant from a staunch Protestant family. He was not reared in the sceptical Court circles—he did not imbibe his wisdom from the heirs of Montaigne—nor was he reared in the centers of Cartesianism and anti-Cartesianism. Rather he was nourished in the controversies between the hounded Calvinists and the triumphant Catholics.⁸⁰ At the Jesuit college in Toulouse, he learned a style of argumentation that was first to convert him and then transform him into a "Protestant in the full sense of the word, one who is against everything that is said and everything that is done."⁸¹ For a brief period, Bayle became a Catholic because of arguments showing that the Protestants had no certain basis for their position. Having argued himself into Catholicism, he then argued himself back into Protestantism. He next studied at Geneva, taught philosophy at Sedan and at Rotterdam, and then devoted himself almost entirely to controversies and scholarship, culminating in the great *Historical and Critical Dictionary* of 1697, revised and expanded in 1702. He died, pen in hand, in 1706 answering

another of his many opponents. Since the *Dictionary* is so enormous (seven to eight million words), it is somewhat difficult to see what the author is trying to accomplish besides attacking everyone and presenting his salacious interpretation of the great events of human and divine history. Having recently read it all from Aaron to Zuylichem, plus all of the appendices and supplements, I am now convinced that it is all of a piece, that it all aims at a central message, and that it is really a drama, with its heroes and villains turning up in the folio columns of notes and notes to notes. In the brief time available, I shall try to sketch my interpretation in terms of three of Bayle's heroes, the "subtle" Arriaga, Moses Maimonides, and Pierre Buel.

The "subtle" Arriaga, the last of the great Spanish Scholastics, who died in 1667, Bayle tells us was accused of being a Pyrrhonist because he was so good at destroying other theories, and so poor at defending the one he was supposed to.⁸² Bayle tackles intellectual issues in the style of the "subtle" Arriaga, rather than in the style of the heirs of Montaigne. Every theory in any area whatsoever is of interest, not just Cartesian theories. And, in fact, Bayle exhibits a clearer preference for metaphysical theories than for epistemological ones. Each theory is inspected and examined and questioned, and in the course of this process, it disintegrates into contradictions and paradoxes. Recently, Madame Labrousse has discovered that the pattern of the approach appeared in the lectures Bayle attended at Toulouse (given no doubt by a spiritual relation of the "subtle" Arriaga). This approach appears again in Bayle's lectures at Rotterdam, and over and over again in the *Dictionary*. It takes the form of degenerating various theories into versions of the Zeno paradoxes, impossible dilemmas, etc. Perhaps, one form of it can be gleaned not from the magnificent notes to the articles on Zeno and Leucippus, etc., but from Leibniz's description in the *Theodicy* of what an argument with his friend Bayle was like. Leibniz reports that if one asserted something, Bayle would proceed to analyze the assertion, and question it, until one was ready to give up and assert

⁸⁰ For details about Bayle's life and especially his Protestant background, see the new excellent biography by Elisabeth Labrousse, *Pierre Bayle, Tome I, Du pays de Foix à la cité d'Erasmus, International Archives of the History of Ideas*, vol. 1 (The Hague, 1963).

⁸¹ Cardinal de Polignac reports Bayle told him, "Oui, Monsieur, je suis bon Protestant, & dans toute la force du mot; car au fond de mon ame, je proteste contre tout ce qui se dit & tout ce qui se fait." Cited in De Boze's "Eloge de M. le Cardinal de Polignac," in de Polignac's *L'Anti-Lucretius*, Tome I (Paris, 1749), p. 13.

⁸² Bayle, *Dictionnaire*, art. *Arriaga* (Roderic de), Rem. B and C.

its opposite, and Bayle would then proceed to analyze *this* assertion and question it, and so on. This critical analysis, Leibniz reported, never ended as long as the opponent remained present.³³ The Earl of Shaftesbury, who had lived under the same roof with Bayle, said, "Whatever opinion of mine stood not the test of his piercing reason, I learned by degree either to discard as frivolous, or not to rely on with that boldness as before; but that which bore the trial I prized as purest gold."³⁴ And Bayle, unlike the other sceptics, has no favorite or preferred target. Descartes, Spinoza, Malebranche, Aristotle, Epicurus, Gassendi, or anyone else was fair game. As soon as the "incomparable" Isaac Newton turned up, Bayle attacked him. When his friend John Locke appeared in print, Bayle proceeded to work on his theories, and so on.³⁵ And, in the article *Rorarius*, where Bayle expressed his delight with the efforts of his friend, Gottfried Wilhelm von Leibniz, it was mainly because the German genius had developed a new theory, that of the pre-established harmony, that could give Bayle a new chance for destructive analysis.³⁶

In each case, Bayle is not solely or merely concerned to challenge a theory, but to use the occasion to generalize the attack to all theories, and to show the hopeless abysses to which all human intellectual endeavors lead. When, in article *Pyrrho*, remark B, he starts with the Cartesian theory, he quickly generalizes the critiques of the previous sceptics into an attack on the entire rational world and raises the horrendous possibility, that *no* previous sceptic had entertained, that a proposition could be self-evident, and yet demonstrably false, that there might be no criterion of truth whatsoever. Up and down the folio columns, Bayle plays the role of the "subtle" Arriaga, attacking, destroying, and dissolving. Basically his *method* is the same as that of Leibniz in analyzing problems. Their texts read much the

same when they start on a consideration of some one else's theory. They proceed as if they were peeling an onion, tearing off layer after layer of contradiction and absurdity. They both love to do this, especially with metaphysical theories about the nature of matter, or motion, or the relation of the soul and the body or theological theories about the nature of evil or transubstantiation. Leibniz claimed that by starting with the doubts of Sextus Empiricus, and this kind of sceptical analysis, one would end up with the essential kernel of truth ("des belles principes") from which to reconstruct the "true" rational world.³⁷ For Bayle, the method works as well on Leibniz as on anyone else. The onion is peeled until *nothing* is left.

As the destruction and dissolution of theories proceeds, Bayle keeps pointing out the lesson to be learned from this and shifts from Arriaga to Maimonides (whom he cites an amazing number of times throughout the *Dictionary*). Every investigation of any rational claim, be it in philosophy, theology, or science, leads to perplexities. The people who try to proceed in a rational or scientific manner are reduced to being perplexed, being "les égarés." Reason is "la voie d'égarements." When this point, which is drilled home over and over again in these terms, is realized, Bayle insists one will abandon rationality as a guide and seek a new guide.³⁸ The terms "égarés" and "guide" are used with such amazing frequency that they suggest Bayle was proposing the *Dictionary* as a new Maimonidean work, a new Guide for the Perplexed.³⁹

In the role of the "subtle" Arriaga, Bayle provides the high road to Pyrrhonism, the way by which each and every rational man becomes perplexed and seeks another guide. This high road to Pyrrhonism is the utter and total *reductio ad absurdum* (quite literally) of all of our intellectual pretensions. Thus, the *Dictionary* is really two guides—one for becoming completely perplexed,

³³ Gottfried Wilhelm von Leibniz, *Essais de Theodice sur la bonté de Dieu, la liberté de l'homme et l'origine du mal* (Amsterdam, 1710), p. 547.

³⁴ Anthony Ashley Cooper, Third Earl of Shaftesbury, letter to Mons. Basnage, Jan. 21, 1706(7), in *The Life, Unpublished Letters, and Philosophical Regimen of Anthony, Earl of Shaftesbury*, ed. by Benjamin Rand (London, 1900), p. 374.

³⁵ On Newton, see Bayle's articles *Leucippus*, Rem. G, and *Zeno of Elea*, Rem. I. Locke is discussed in several articles. See especially *Dicaearchus*, Rem. M, *Perrot*, Rem. L, *Rorarius*, Rem. K, and *Zeno of Elea*, Rem. I.

³⁶ Bayle, *Dictionnaire*, art. *Rorarius*, Rem. H and L.

³⁷ Cf. Leibniz' letter to the abbé Nicaise on the death of Simon Foucher, cited in Rabbe, *L'Abbé Simon Foucher*, p. 19. See also Leibniz' letters to Foucher of Jan. 1692 (an extract of this appeared in the *Journal des Sçavans*), in *Die Philosophische Schriften von Gott. Wilh. Leibniz*, Hrsg. von C. J. Gerhardt, vol. 1 (Berlin, 1875), p. 402.

³⁸ See, for example articles *Bunel* (Pierre), Rem. E, *Manicheans*, Rem. D, *Pomponazzi*, Rem. G, *Pyrrho*, Rem. B and C, and the Clarifications on Atheism, Manicheanism, and Pyrrhonism added to the *Dictionary*.

³⁹ The title of Munk's translation, of much later vintage, of course, is *Le Guide pour les égarés*. Bayle read Maimonides in Latin, and always gives the title as a transliteration of the Hebrew, *Morsh Nebochim*.

by trying to make sense of any aspect of our universe, and the other for overcoming complete perplexity, by turning to faith and Revelation.

Bayle, in the role of the new Maimonides, keeps urging faith and Revelation as the only way of answering any problems. In surveying the need for faith from this perspective, Bayle points out that the rational world—any and all of man's intellectual pretensions—is built on maxims which are *évident* (self-evident) but false. All science is built on the maxim, *ex nihilo nihil fit*, which is demolished by Genesis 1: 1. Morality is built on maxims that are 'shattered by reading the lives of the Patriarchs, the heroes of ancient Israel, and the lives of the saints.⁴⁰ Philosophy and theology are built on maxims that are incompatible with the basic truths of Christianity, such as the doctrine of the Trinity or of the Fall.⁴¹ The rational man can start off with the best intentions to comprehend an aspect of man's moral, scientific, philosophical, or theological cosmos. As long as he remains within the bounds of rationality, employs the tools of reason, its basic maxims, etc., he can only construct theories that are "big with contradiction and absurdity."⁴² When he gives up in complete perplexity, and turns to faith and Revelation, he finds out the explanation of this sad state of affairs, namely, that the Good Book reveals that God made and runs a radically and totally different world from that which the rational man starts with and reasons in.

Throughout the *Dictionary*, Bayle is always claiming that his scepticism is a preparation for the faith. The appendices, written because his local church, the French Reformed Church of Rotterdam, had its doubts after reading articles *David*, *Manicheans*, *Pyrrho*, and all the smutty stories in the *Dictionary*, make this the central claim of the work.⁴³ And it is this claim that Bayle defends against all-

comers thereafter, be they fanatic orthodox types like Jurieu, or woolly-minded liberals like Le Clerc, Jaquelot, or Leibniz.⁴⁴

Every foray into the rational world is turned into a way of bringing home the need for a *Guide for the Perplexed*, and an occasion for the fideistic message. The magnificent clarifications, especially the third one on Pyrrhonism, spell out the conflict of reason and faith in almost Kierkegaardian terms, and always advocate the acceptance of faith without reason. The ship of Jesus Christ was not made to sail on rationalist seas. The best faith, we are told, is that which is built on the ruins of reason. The tedious *Réponse aux questions d'un provincial*, and the final opus, *Entretiens de Maxime et de Themiste* keep stating this view.

In the course of developing this case, Bayle extends every type of classical sceptical argument, and every imaginable one, into a total dissolution of all of man's intellectual pretensions. He leaves nothing alone, and is happy to show that all theorizing, regardless of subject matter, is incoherent, inadequate, and inconsistent. The new science, and the newest science (of Newton) fare no better than Scholastic views (which Bayle claimed have only the virtue that when the parents come to hear their scholar-son dispute, the son, instead of just being silent before insoluble difficulties, can rattle off the sentences full of meaningless jargon. This makes the parents happy, though it answers nothing.)⁴⁵

Only irrational faith is left. The Revelation that states the content of this faith shows that God made a world which makes no sense to us, and runs it in a completely unintelligible manner. Those who try to measure this strange Divine Universe by human standards commit the dreadful Socinian heresy, that reason is and ought to be the rule of truth.⁴⁶ If they try to maintain this heretical

⁴⁰ Bayle's article *David* is the notorious example. Many of his other articles on Biblical characters, such as *Abimelech*, *Ham*, and *Sarah* suggest the same point. His articles on St. Francis of Assisi, St. Mary of Egypt do this also.

On the famous article *David*, see the excellent recent study by Walter Rex, "Pierre Bayle: The Theology and Politics of the Article on David," *Bibliothèque d'Humanisme et Renaissance*, vols. 24-25 (1962-63), pp. 168-89 and 366-403.

⁴¹ Article *Pyrrho*, Rem. B, develops this point at length.

⁴² This is David Hume's phrase, from his *Enquiry Concerning Human Understanding*, § xii, pt. ii, Selby-Bigge ed. (Oxford, 1951), p. 157.

⁴³ The criticisms by the Rotterdam Consistory of the French Reformed Church and Bayle's answers appear in "Actes du Consistoire de l'Eglise Wallone de Rotterdam, concernant le Dictionnaire Historique et Critique de Mr. Bayle," which is published in Tome I of the *Dictionnaire* 5^e edition, 1740, pp. cxv-cxx. The first three appendices on atheism, Manichaeism, and Pyrrhonism develop Bayle's "fideism," and the last gives the strongest statement of Bayle's view.

⁴⁴ Cf. Bayle's major works written after the *Dictionary*, *Réponse aux questions d'un provincial*, and *Entretiens de Maxime et de Themiste*, republished in Bayle's *Oeuvres diverses*, Tomes III and IV (La Haye, 1727).

⁴⁵ Bayle, *Dictionary*, art. *Zeno of Elea*, Rem. G.

⁴⁶ Cf. Bayle, *Dictionary*, art. *Socinus* (Faustus), at the end; and "Suite des réflexions sur le prétendu jugement du public," § XXVIII. (This essay is usually published in the supplements to the *Dictionary*.)

outlook, they will only end up in complete bewilderment and perplexity. Then they will be ready, finally, for *The Guide for the Perplexed*—faith and Revelation.

In spite of the myriad number of times Bayle stated his fideism, his opponents were sure that his point was not the destruction of reason for the sake of religion, but rather the reverse—the destruction of religion for the sake of reason. The blatant impiety of Bayle's salacious discussions of Old Testament heroes, Church Fathers, and saints, and the *reductio ad absurdum* of all theologies, be they Christian or pagan, certainly make one wonder what the real point may be. Historically, I think it is worth noting that Bayle remained a member of the small and contentious French Reformed Church of Rotterdam to the end of his life, rather than leaving it for the growing community of non-believers, or for one of the many liberal churches then extant in Holland. He fought furiously against any liberals and/or Unitarians who thought he was their ally, and he defended himself tooth and nail before his own consistory, always insisting he was a true follower of John Calvin.

Was he himself misguided, or was he just joking? The biographical facts in the case seem to make the latter alternative unlikely, since it would require that Bayle devote all his energies and all his time to a fairly silly practical joke that only a few insightful cohorts in the French Reformed Church might appreciate. If he were genuinely fighting the war of science against religion, it would seem that he could well have carried the battle on in the larger Republic of Letters, without bothering with his brethren in the French Reformed Church, and without going to such odd

and elaborate ways of joking them. The weapons used, if he were joking, were so destructive against rational, scientific activities, that it is hard to see how he might have intended to aid in a victory of reason over faith.⁴⁷

Bayle, it seems to me, provides an answer to what he has in mind in his article on his third hero, Pierre Bunel. In the 16th century Bunel was an obscure pedant from Toulouse who accidentally triggered off the development of modern scepticism by giving Montaigne's father a copy of Raimond Sebond's *Natural Theology*, and thereby provided the occasion for Montaigne's translation of Sebond, and his *Apology for Raimond Sebond*, the greatest classic of modern scepticism. Bunel is presented by Bayle as a *true* Christian, whose Christianity consisted in a quiet life devoted exclusively to scholarly pursuits.⁴⁸ No indication is given that he suffered from the immoral passions of the Patriarchs or the saints, or that he suffered from the madness of the Church Fathers or the religious mystics. Bunel, according to Bayle, lived a life very much like that of Pierre Bayle. And, Shaftesbury, who knew Bayle very well, called him "one of the best of Christians" for approximately the same reasons that Bayle said this of Bunel.⁴⁹ The Erasmian life of pious study is made the essence of the Christian life.

For many people, the life and works of both Bayle and Bunel would hardly seem to be striking examples of Christian behavior, in good part, I think, because they appear to be devoid of all religious concern and passion. Bayle, in his most fideistic passages, may read like Pascal or Kierkegaard, but the flavor and feeling of real religious spirit definitely seems to be lacking.⁵⁰ Bayle's appeals to the *Bible* also suggest this lack. He

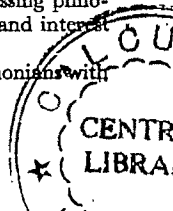
⁴⁷ In recent years, several commentators have expressed the view, in various ways, that Bayle was genuinely or seriously a believer. Among these are W. H. Barber, of the University of London, H. M. Bracken, of Arizona State University, C. Brush, of Columbia, P. Dibon, of Amsterdam, E. Labrousse, of Paris, W. Rex, of University of California, Berkeley, and K. Sandberg, University of Arizona. Barber's article, "Pierre Bayle: Faith and Reason," in *The French Mind: Studies in Honor of Gustave Rudler* (Oxford, 1952); plus the articles by Dibon, Labrousse, and myself in *Pierre Bayle, Le Philosophe de Rotterdam*, ed. by P. Dibon (Amsterdam, 1959); plus Madame Labrousse's prodigious researches on Bayle, have stimulated a serious reconsideration of the usual interpretation of Bayle as the precursor of the irreligiosity of the Enlightenment, and have led others to take seriously the possibility that Bayle was a sincere Calvinist of some sort.

Recently two articles in *French Studies*, E. D. James's "Scepticism and Fideism in Bayle's *Dictionnaire*," vol. 16 (1962), pp. 308-323, and H. T. Mason's "Pierre Bayle's Religious Views," vol. 17 (1963), pp. 205-217, have raised serious challenges to the religious interpretation of Bayle.

⁴⁸ Bayle, *Dictionary*, art. *Bunel* (Pierre), exp. Rem. C.

⁴⁹ Shaftesbury, letter to Mr. Darby, Feb. 2, 1708, in *Life, Unpublished Letters and Philosophical Regimen of Anthony, Earl of Shaftesbury*, pp. 385-386, "he was in practice one of the best of Christians, and almost the only man I ever knew who, professing philosophy, lived truly as a *philosopher*; with that innocence, virtue, temperance, humility, and contempt of the world and interest which might be called exemplary."

⁵⁰ Compare, for example, Bayle's art. *Pyrrho*, Rem. C, with Pascal's *Pensée* no. 494; or the clarification on the Pyrrhonians with Kierkegaard's *Philosophical Fragments*, *Concluding Unscientific Postscript*, or *Training in Christianity*.



always regards the *Old Testament* as fair game for Baylean analysis at its highest, and he revels in figuring out the most immoral and ridiculous possibilities about Adam and Eve, Abraham and Sarah, what went on in Noah's Ark, and so on. The *New Testament* figures are not discussed. There are no articles on such obvious Baylean topics, such as the conversion of St. Mary Magdalene (though he has a delightful, racy one on a later and very similar case, St. Mary of Egypt).⁵¹ Jesus is not presented very often as a major moral or religious figure. The main appeal to the *New Testament* is solely for the extreme fideism and anti-rationalism of St. Paul's *1st Corinthians*, along with some other similar lines culled hither and yon. I doubt that Bayle omitted the events in the *New Testament* out of fear of criticism since he certainly did little else to avoid giving his opponents a chance to complain. His one-time mentor and ally and later fiercest enemy, Pierre Jurieu, had insisted that all of the Bible must seem odd to people without grace and had offered pretty shocking examples, such as how the story of the Virgin Birth might appear to unbelievers.⁵² Bayle could easily have embellished these possible infidel readings. Instead he chose to say nothing about the message and meaning of the *New Testament*.

A possible explanation of Bayle's own religious views appears in remark M of article *Spinoza*, where, after stating the fideistic message once more, Bayle describes two religious conditions, one of those who have religion in the heart and not in the mind, and the other of those who have religion in the mind, but not in the heart. Bayle may well have been in one of these categories, and hence,

constantly devoting himself to destroying reason for faith, but unable to state the faith with any feeling or fervor.⁵³ He could work out the logic of religious belief and show that it involved an overall scepticism about everything in the rational world. He could piously study, dissect, and dissolve all rational and scientific pretensions. In so doing, he suffered no *angst* or despair or fear and trembling since religion was only in his mind. His inability to give more than a tepid statement of the Judeo-Christian message, then, could only give his analysis of man's intellectual state a hollow and ludicrous air, since it lacked the emotional force of those who had religion in the heart—the Pascals, Hamanns, Kierkegaards, Lamennaises, and Dostoevskis of this world.

Bayle's all-out attack on everything that is said and everything that is done carried scepticism to its ultimate extreme. The other sceptics of the 17th century undermined the Cartesian revolution to prepare the way for a philosophy and a science without metaphysics. Bayle undermined all of man's intellectual efforts and left an incoherent shambles as the legacy of the new philosophy and the new science. He showed that all approaches to all problems soon dissolve into meaninglessness and incoherence. The problem was not just the inner structure of the Cartesian theory of knowledge, but the basic irrationality of God's world, that exhibited itself in all efforts at theorizing. Bayle's method was not just that of the other sceptics, the various tropes of Sextus Empiricus, but it was a method of analysis, in the style of the "subtle" Arriaga, and of Leibniz, a method which in Bayle's hands, led only to utter confusion.

⁵¹ In fact, there is only one article on anyone in the *New Testament*, namely the one on St. John the Evangelist. There are a great many on Old Testament figures and on later "heroes" of Judaism and Christianity.

⁵² Pierre Jurieu's theory about the problem of reading the Bible without Grace is developed in his *Le Vray système de l'église & la véritable analyse de la foy* (Dortrecht, 1686), chaps. xii-xiii; and *Seconde Apologie pour M. Jurieu, ou réponse à un libelle sans nom présenté aux Synodes de Leyden & de Naerden*, . . . (Rotterdam, 1692).

Jurieu's examples of his theory were regarded as shocking by many of his contemporaries.

⁵³ Which category to put him in depends on how one interprets the text in question from art. *Spinoza*, Rem. M. The text states "The Abbé de Dangeau speaks of certain people who have religion in their minds, but not in their hearts. They are convinced of its truth, without their consciences being affected by the love of God. I believe that one can also say that there are people who have religion in their hearts, but not in their minds. They lose sight of it as soon as they seek it by the methods of human reasoning. They do not know where they are while they compare the pro and con. But as soon as they no longer dispute, and as soon as they listen only to the proofs of feeling, the instincts of conscience, the weight of education, etc., they are convinced of a religion, and they conform their lives to it as much as human weakness permits. Cicero was like this. We can hardly doubt this when we compare his other books with that of the *De Natura Deorum* where he makes Cotta (the sceptic) triumph over the other interlocutors who maintained that there are Gods."

Bayle's completely unemotional statements of the faith would indicate he did not have "religion in the heart" (at least in Pascal's sense). His description of Cicero, however, suggests that Bayle and he might fall in the same category. "Religion in the mind," however, suggests more of the fervorless flavor of Bayle. In a recent discussion with Madame Labrousse, we agreed that this passage was crucial for characterizing Bayle's own religious views, though we could not agree as to whether to characterize Bayle as having religion in the heart, or in the mind. In this paper, I have used the metaphor "religion in the mind" to indicate the unemotional character of Bayle's religion.

bewilderment, and perplexity, not to an admiration of empirical study. Every problem and every theory, Bayle showed, ended in contradictions, absurdities, and paradoxes. Only a new guide—faith and Revelation could lead man out of this morass. But Bayle, like Bunel, could only find this faith in the mind, and could only exhibit it in a quiet life of interminable scholarly endeavor, instead of a life of entire religious commitment, like Kierkegaard's Knight of Faith.

In terms of this picture, Bayle has less in common with his Enlightenment heirs than do the other 17th century sceptics. Gassendi, Sorbière, Foucher, and Huet may actually have supplied the rationale for a science without metaphysics that was to answer all of man's problems by destroying the Cartesian enterprise of a science based on metaphysical knowledge. Bayle, while providing the Arsenal of the Enlightenment, the weapons, and the ammunition that were to be fired at all of the opponents of the Age of Reason, had no illusions, himself, about what man's reason could accomplish. David Hume, alone, of the major Enlightenment figures, seems to me to have remained in Bayle's shadow, to have worried about the specific problems Bayle posed, and to have found all efforts, including his own, to resolve these difficulties utterly fruitless. Reason, he found, always undermines itself when it tries to find consistent and coherent answers.

After one such discussion, Hume could only conclude,

This sceptical doubt, both in regard to reason and the senses, is a malady, which can never be radically cur'd, but must return upon us every moment, however we may chace it away, and sometimes may seem entirely free from it. 'Tis impossible upon any system to defend either our understanding or senses; and we but expose them farther when we endeavour to justify them in that manner. As the sceptical doubt arises naturally from a profound and intense reflection on those subjects, it always encreases, the farther we

carry our reflections, whether in opposition or conformity to it. Carelessness and in-attention alone can afford us any remedy.⁵⁴

Hume could recommend this careless scepticism at various times and suggest that when one became too overwhelmed with doubts it was better to leave the philosophical cabinet, to act as any ordinary man would, and to engage in diversions such as backgammon rather than further destructive sceptical analysis.⁵⁵ At the end of his *Dialogue concerning Natural Religion*, he could even suggest that "To be a philosophic sceptic is, in a man of letters, the first and most essential towards being a sound, believing Christian."⁵⁶ But, since Hume had no religion in the heart, and probably even less in mind than Bayle had, he could only advance such a pallid deism that it could serve no one as a support against the waves of the new dogmatism of the Enlightenment. "The wise in every age conclude," as Blacklock said, "what Pyrrho taught and Hume renewed, that dogmatists are fools."⁵⁷

The only solution Hume could offer, basically, was the benevolent, irrational actions of Nature. "Philosophy would render us completely Pyrrhonian, were not Nature too strong for it."⁵⁸ Nature, acting by "an absolute and uncontrollable necessity" saves us from going stark raving mad in the face of the flood of doubts with which Bayle had inundated the 18th century.⁵⁹ Bayle's odd trinity of forces that account for our beliefs, "Grace, education or prejudice,"⁶⁰ or Hume's benevolent irrational Nature, were all that were left to keep man afloat in the seas of uncertainty. The sceptical crisis that the new philosophy was designed to resolve was to become our permanent heritage. The Enlightenment heroes might think that a faith in reason and science would solve all. But Bayle had already undermined this faith and offered a blind, meaningless one instead. Hume, following in his footsteps, was to offer an animal faith as the only way we do in fact survive intellectually in a

⁵⁴ David Hume, *A Treatise of Human Nature*, Selby-Bigge ed. (Oxford, 1951), p. 218. For a lengthier discussion of Hume's relations to Bayle, see R. H. Popkin, "Scepticism in the Enlightenment," in *Transactions of the 1st International Congress on the Enlightenment*, Geneva, July, 1963, pp. 1321-1345, and "Bayle and Hume," *Transactions of the XIIIth International Congress of Philosophy*, Mexico City, September, 1963, forthcoming.

⁵⁵ Hume, *Treatise*, Bk. I, Pt. IV, § vii, "Conclusion" of this book.

⁵⁶ David Hume, *Dialogues Concerning Natural Religion*, Kemp Smith ed. (Edinburgh, 1947), p. 228.

⁵⁷ Original version of a poem by Thomas Blacklock, as it appeared in Hume's letter of April 20, 1756, to John Clephane, in *The Letters of David Hume*, ed. by J. Y. T. Grieg, vol. 1 (Oxford, 1932), p. 231.

⁵⁸ Hume, *An Abstract of a Treatise of Human Nature* (Cambridge, 1938), p. 24.

⁵⁹ Cf. Hume, *Treatise*, Pt. IV, § 1, pp. 183-84.

⁶⁰ Bayle uses this list in art. *Pyrrho*, Rem. B, *Simonides*, Rem. F, and elsewhere. He discusses the differing effects of education and Grace in art. *Pellison*, Rem. D.

world in which our rational world is actually naught but "a sink of uncertainty and error." Others might exude confidence and optimism in the Age of Reason, but Hume was to appreciate Bayle's discordant note that was to resound throughout the centuries to come, while men sought for a solution from uncertainty in a new religion of the heart, or in a new form of animal faith, as each new dogmatism fell victim to the unlimited and unlimitable doubts that Bayle had raised.

University of California, San Diego
La Jolla, California

III. PRESUPPOSITION AND ENTAILMENT

G. NERLICH

I. INTRODUCTION

THE purpose of this paper is to examine the logical, or quasi-logical, relation of presupposition. The idea of one proposition's presupposing another has had in the recent past, and continues to have, considerable currency and it is appealed to in a wide range of philosophical questions. In the form in which it is most often invoked, the only reasonably determinate and worked out account of it is due to Mr. P. F. Strawson and is given in three places. ("On Referring," *Mind* [1950]; reprinted in *Essays in Conceptual Analysis*, ed. A. G. N. Flew. Page references to this paper will here refer to the Flew volume. P. F. Strawson, *Introduction to Logical Theory*, pp. 174-179. "Presupposing: a Reply to Mr. Sellars," *Philosophical Review* [1954], pp. 216-231. I shall use the abbreviations "Referring," "Introduction," and "Presupposing" to refer to these respectively.) I wish to argue that Strawson's account is in various ways confused or mistaken. Consequently, I claim that we do not have any clear useful notion of presupposition as a *distinct* logical relation, though my sole argument on behalf of this claim is given in the following criticisms of Strawson's account.

1. The current (Strawsonian) doctrine of presupposition.

The interest of the notion of presupposition lies in the claim that while we can, in some sense, argue from one proposition to another which it presupposes, yet the one does not *entail* the other.¹ It appears that a new logical relation between propositions has been discovered, a new form of

implication which Strawson says is certainly not logical implication (*Referring*, p. 34). Such a discovery would be, quite obviously, of great philosophical and logical importance.

It is not possible, at present, to be more explicit about the Strawsonian doctrine, for reasons which will be given in Section II.

2. Theses to be defended in this paper.

I shall argue (against Strawson) for the following main conclusions: (1) The doctrine oscillates between two quite distinct, indeed incompatible, positions. (2) The first of these, though it has some plausibility in itself, will not sustain the case that a new logical relation has been discovered. (3) The second of these positions is false.

The bulk of the paper is directed at arguing for thesis (2), which requires quite detailed treatment. It is dealt with in two parts (2a) and (2b) in Sections III and IV.

II. THESIS (1)

The Doctrine Oscillates between Two Incompatible Positions.

3. The two positions.

My first thesis is that, taking the three papers as a whole, it is *unclear* whether there is only one or whether there are two positions, which Strawson actually advances. If this thesis can be sustained, as I believe it can, then his account suffers from a major obscurity. I will now outline the two positions which it seems to me are each suggested by a part of what Strawson says. It will be quite

¹ It is *this* notion of presupposition with which the present paper is concerned. It is, of course, quite possible to give a clear-cut sense to a logical relation "presupposition" where this is construed as a special form of entailment. See N. Rescher, "On the Logic of Presuppositions," *Philosophy and Phenomenological Research*, vol. 21, no. 4 (June 1961), and J. Hintikka, "Existential Presuppositions and Existential Commitments," *Journal of Philosophy*, vol. 56, no. 3 (January 29, 1959). Nor are such attempts new. See J. P. N. Land, "Brentano's Logical Investigations," *Mind*, vol. 1 (1876), in which Land uses the word "presupposes" to express the relation between, e.g., a singular categorical and its related existential proposition. Here it is clear, however, that Land's presupposition is again a special case of entailment. A view which is essentially equivalent to Strawson's and which distinguishes presupposition from entailment was put forward, also in 1950, by Mr. P. T. Geach (see his "Russell's Theory of Descriptions," *Analysis*, vol. 10, no. 4 [1950], reprinted in M. Macdonald, *Philosophy and Analysis* [Oxford, 1954]). According to Geach, if the presupposed statement is false then the presupposing statement is "out of place." There is no indication exactly what Geach means by the quoted phrase except that an out of place proposition has no truth value.

clear, I think, that the positions are incompatible.

The first position is the following one: *when a meaningful sentence is uttered by a speaker of the language on a certain occasion, then a necessary condition for the statement thus made to have a truth value is that the (main) referring expression in the sentence has, on that occasion, a reference.* The second position is this one: *when a meaningful sentence is uttered by a speaker of the language on a certain occasion, then a necessary condition of the speaker's thereby making a statement is that the (main) referring expression in the sentence has, on that occasion, a reference.* It is to be noticed that the second position denies that there is any statement while the first asserts that there is one but denies only that it has a truth value, when failure of reference occurs. It is obvious that the relation of presupposition which is in question here will be very different depending on which of these positions is adopted.

4. Evidence for the occurrence of both positions in Strawson's accounts.

Strawson commits himself quite explicitly to the first position in both the later treatments of the subject; in *Introduction* (p. 175) and in *Presupposing* (pp. 216-217), where he refers again to *Introduction* and *passim*. There can be no question at all as to the correctness of attributing that view to him. The contentious issue is whether, in the first paper, *Referring*, and in some remarks in the later two, he is also advancing the second position. Certainly it cannot be said that he is there *explicitly* committed to the second position nor does he say anything *explicitly* inconsistent with the first position: indeed, as Sellars pointed out, *Referring* puts forward *no* explicit clear-cut view as to what presupposition is.

In answering the question whether Strawson takes up the second position one cannot be nearly so direct or certain. To be sure, there are odd remarks made in this connection, such as "And not all general sentences which look as if they are used to make statements are so used" (*Introduction*, p. 195). But this is uncertain evidence. Mainly I rest the case on Strawson's emphasizing the role of the sentence-statement distinction: I cannot see how this can be relevant to the first position, as I shall argue. On the other hand I do not think we can base the second position on the distinction either. Nor am I clear as to which of the considerations mentioned by Strawson in dealing with the distinction are the ones he would think

establish that the sentence "The king of France is bald" fails to make a statement if uttered by a speaker of English in 1965. However, it is to be noticed that my failure to find any real relevance in Strawson's strong emphasis on the sentence-statement distinction creates some presumption that I misinterpret him on the point. I hope that my discussion will be detailed enough to make the presumption small and in any case the precise definitions of presupposition are certainly as I have stated.

What really makes this second position worth discussion is not the fact that it is clearly there in Strawson—for that is decidedly obscure—but the fact that a number of professional philosophers have taken it for granted, presumably on the basis of *Referring*, that this is indeed Strawson's main view.² These included several of my colleagues and myself, and perhaps my colleagues are still unconvinced. However, I shall give some reasons for thinking that Strawson does adopt this position, though the list of reasons is not exhaustive.

(a) If Strawson's position were just the first one, then he must hold that Russell's essential mistake is just that he supposed the law of excluded middle to hold for propositions or statements. Russell's view on sentences like "The king of France is bald," or something essentially equivalent to his view, as it concerns entailment of existential propositions follows directly from these theses: (i) every statement is true or false; (ii) a statement is made by uttering the meaningful sentence "The king of France is bald" on an occasion when there is no king of France; (iii) no true statement is made on such an occasion. At no stage does Strawson begin to deny (iii). Now his first position is compatible with, and indeed entails, thesis (ii). On this interpretation, then, it is possible for Strawson to disagree with Russell only on thesis (i). I cannot see at all how it can be maintained that Russell's or any other's belief in (i) can involve a confusion between sentence and statement, or the bogus trichotomy "true, false or meaningless." Nor do I feel my own trust in thesis (i) at all undermined by its being pointed out that it is improper to insist on an analogue of thesis (i) for sentences. In short, the most prominent feature of *Referring*, viz., the absolutely decisive arguments against confusing sentence and statement, neither undermine Russell's position nor support Strawson's if it is being conceded from the beginning that in the relevant cases statements are made. For then any confusion

² See R. Hancock, "Presupposition," *Philosophical Quarterly*, vol. 10, no. 38 (1960).

there may have been in Russell between sentence and statement is quite irrelevant to the consideration of the statement that the king of France is bald.

On the other hand, all this very prominent argumentation is central if Strawson's position is the second one. For then, he is denying thesis (ii) above, to which confusions between sentence and statement might very well be entirely crucial.

(b) In *Referring* and *Presupposing*, Strawson produces cases about which we are indeed tempted to deny that there are statements. There are two kinds of cases; first, fictional utterances which are explicitly disavowed as statements in *Presupposing* (p. 221) and closely assimilated to the cases relevant to Russell's analysis in *Referring* (p. 35); second the example is given of uttering the sentence "This is a fine red one" in circumstances where there is nothing in the cupped hands offered for inspection by the speaker. It is reasonable to assume that the utterance of "The king of France is bald" is being similarly treated. What else is the point of examples to show that meaningful indicative sentences may not make statements when uttered on certain occasions?

(c) Throughout *Referring* Strawson treats referring expressions and sentences quite analogously. He says that a significant referring expression may fail to refer (or its utterer may fail to refer). The analogous thing about sentences would *seem* to be that the utterer fails to state anything. It must be conceded that Strawson uses the expression "fail to make a true-or-false assertion" throughout. But if this is the first position then the analogy between sentences and referring expressions seems to be imperfect, and hence misleading. For the sentence (or its utterer) *does* succeed in asserting something, though it happens to be neither true nor false.

On this basis it seems reasonable to suppose that Strawson *has* confused the first and second positions, which are, in fact, incompatible. But in any case it seems quite safe to conclude that considerable obscurity surrounds the question what relevance the sentence-statement distinction has to Strawson's case against Russell.

III. THESIS (2a)

The First Position does not require that presupposition be distinguished from entailment. Moreover a denial that the former is a special case of the latter can be sustained only by legislating on the meaning of logical terms.

5. *What the relation of presupposition is, given the first position.*

The definitive statement of the first position is to be found in *Introduction* (p. 175) and to this place the reader of *Presupposing* is again referred. It is clear from this that Strawson commits himself to a denial of the law of excluded middle in the form in which it states that every proposition (statement) is either true or false. It is also clear from a number of passages in *Presupposing* (e.g., pp. 224-225). The position, to put it briefly, is this: If one proposition is a necessary condition both of the truth *and* the falsity of another, then the second is said not to entail but to presuppose the first; and the conjunction of the negation of the presupposed proposition with the presupposing one is said to be not a contradiction but an absurdity of a different kind.

As a further gloss on this, it is clear that Strawson contends that the negation (denial) of "The king of France is bald" is "The king of France is not bald" (*Presupposing*, p. 228); and that the negation (denial) of the classical particular form "SiP" is "SeP," etc. All these propositions *presuppose* existential ones. I shall adopt the following notation: Let 'S' represent a proposition. Then let 'S*' represent the proposition which Strawson wishes to say is presupposed by S and let '—S' represent the proposition which Strawson wishes to call its negation (denial). When S is false, in Strawsonian terminology, then —S is true. But both S and —S may lack a truth value. I shall use '¬S' in the sense defined for it in standard propositional logic, in which the law of excluded middle holds. So that ¬S is the proposition which is true if S is not true. (The signs may be read as "bar" and "tilde" respectively.)

I shall call the logic which contains Strawson's rules for "true," "false," and "entails" the system L_s , and Russellian logic the system L_r . I do not intend to treat them in any rigorous formalized way. I am merely adopting a notational convenience. It is to be noted that '—' is a symbol of L_s whereas '¬' is not, so far as I can see, though there is nothing to suggest that '¬' and '—' differ for existential propositions.

For thesis (2a), I present two arguments for the view that in L_s , S entails S*; they may be called the *a fortiori* argument and the transitivity argument.

6. *The a fortiori argument.*

On the face of it, the concession that S* is a necessary condition both of the truth and the

falsity of S , which Strawson makes in all three papers, is inconsistent with the denial that S entails S^* . For if S^* is a necessary condition of the truth and the falsity of S , it is *a fortiori*, a necessary condition of the truth of S . Or, to put it in another way sometimes favored by Strawson, if S^* is a necessary condition for a question to arise as to the truth or falsity of S then it is a necessary condition, *a fortiori*, for the question to be settled, and *a fortiori* again, to be settled in favor of the truth of S . But if S^* is a necessary condition for the truth of S , then S is a sufficient condition for the truth of S^* . At no point does Strawson begin to deny this. We cannot construe the necessary and sufficient conditions concerned on the basis of the material conditional; for, on that basis, any true statement is a necessary condition of S . Nor can it be the case that Strawson seriously intends the conditional to be *causal* or in any other way contingent—a mere fact about our world. Indeed (*Introduction*, p. 175) he speaks of necessary and sufficient conditions in precisely the logical sense in defining the relation of presupposition. His concession, then, is tantamount to S 's being a *logically sufficient* condition for S^* , which means that S entails S^* .

7. The transitivity argument.

As an independent argument for the view that S entails S^* , it is clear from Strawson's definition of presupposition that if S_2 is a necessary condition just of the truth of S_1 , not of its falsity, then S_1 entails S_2 in L_s . Now it is certainly the case, given Strawson's definition, that the proposition S_2 = "There is in 1965 a bald king of France," is a necessary condition of the truth, but not of the falsity of the proposition S_1 = "The king of France in 1965 is bald." So S_1 entails S_2 . However S_2 obviously entails "There is a king of France in 1965" which is just the proposition which Strawson claims is certainly not entailed or logically implied by S_1 . But if this is so, *we plainly must abandon entailment as a transitive relation*, and this is no light matter. That Strawson is committed to abandoning the transitivity of entailment in L_s is a shock for which his discussion of the matter leaves us totally unprepared. I take it to be quite clear that it would be a matter of legislation to abandon it for a logic purporting to be the logic of ordinary language.

8. Rejoinders to these arguments.

There are two rejoinders which might be made in order to discount the conclusions of the preceding arguments. The rejoinders neither attack

the premisses nor attempt to invalidate the arguments, but they rather suggest that other considerations outweigh them in deciding what we should say and do here. The first has a definite textual backing and the second is, I think, a natural suggestion in the light of Strawson's view.

First (*Introduction*, p. 175), the conjunction $S \& \neg S^*$ is not necessarily false, that is, not a contradiction in L_s . For if $\neg S^*$ is true, then S is not false, so that the conjunction $S \& \neg S^*$ will not be false, though it will lack a truth value. Given that the inconsistency of a conjunction occupies an important place in Strawson's logical theory (see *Introduction*, pp. 2-12) this may seem to be a good reason for rejecting the entailment relation as holding between S and S^* in L_s . It would, of course, not invalidate arguments (a) and (b) and this consideration could only be a recommendation that we should prefer over any other the requirement for entailment that the conjunction of an entailing proposition and the negation of the entailed proposition should be a contradiction.

A second consideration, which again could only be advanced as a recommendation that we prefer other requirements over the logical sufficiency and transitivity requirements might be the following: S^* is also a necessary condition of $\neg S$, the L_s "negation" of S . If we accept the proposition that S entails S^* we should be committed to the proposition that the falsity of S^* , or its negation entails the falsity or the negation of S , i.e., $\neg S$, which is what Strawson explicitly wishes to reject in L_s .

9. Replies to the rejoinders.

The second of these rejoinders is certainly confused. While it is the case in a logic which contains the Law of Excluded Middle, e.g., in L_T , that the proposition " S entails S^* " is equivalent to the proposition " $\neg S^*$ entails $\neg S$," it would be mistaken to suppose that in L_s " S entails S^* " is equivalent to " $\neg S^*$ entails $\neg S$." The equivalence of contraposed conditionals depends upon the law of excluded middle which Strawson rejects.

This non-equivalence of contraposed conditionals in L_s is quite clear and general. Consider the case of a proposition S_2 which is a necessary condition just of the truth of a proposition S_1 , but not a necessary condition of the L_s falsity of S_1 . Such a propositional pair are, S_1 : "My father is a Troll" and S_2 : "There are Trolls." The falsity of S_2 is plainly not a sufficient condition for $\neg S_1$. For a necessary condition of $\neg S_1$ is that I have a father.

So here we have a plain case in L_3 where S_1 entails S_2 but the falsity of S_2 does not entail the falsity of S_1 , or $\neg S_2$ does not entail $\neg S_1$.

A similar flaw cripples the first recommendation. For, once again, it is a quite general feature of L_3 , even for those propositions which Strawson says stand in the entailment relation, that the conjunction of an entailing proposition with the negation of the proposition entailed is not a contradiction. It may be contingently false, of course, but then $S \& \neg S^*$ will also be false under the conditions that S^* is true and S is true. However, the conjunction which Strawson requires to be contradictory may also lack a truth value, just as $S \& \neg S^*$ may.

This can be illustrated, once again, by using the examples S_1 and S_2 above in which S_1 entails S_2 according to L_3 rules. The conjunction "My father is a Troll and it is not the case that there are Trolls" lacks a truth value in L_3 under the condition that I have no father. The conjunction is not false under these circumstances and, of course, not necessarily false.

It seems clear, then, that these recommendations that we should discount the *a fortiori* and transitivity arguments in favor of other features of what we ordinarily mean by entailment, cannot be pressed. For all putative entailment relations in L_3 will have to be disallowed on these grounds, and not merely that between S and S^* .³

IV. THESIS (2b)

A necessary condition of justifying the legislation which distinguishes presupposition from entailment, is false.

10. *Strawson claims that logical advantages in interpreting the classical square of opposition accrue if presupposition is distinguished from entailment.*

Now it is clear, from *Introduction* and *Presupposing*, that Strawson's real interest in abandoning the law of excluded middle and denying that S entails S^* is a systematic one. He regards the rules of his system as justified primarily in that they give us essentially the classical Square of Opposition relations between the ordinary language propositions. He says (*Introduction*, p. 178) "The interpretation I propose . . . gives the constants of the system *just the sense* which they have in a vast group of statements of ordinary speech." (My emphasis). Again, in *Presupposing* (p. 228), Strawson mentions features of the logic of ordinary

language which his doctrine of presupposition renders consistent and intelligible. On p. 230 he says "I should represent myself as trying to describe the actual logical features of ordinary speech." The features are the following ones (*Presupposing*, p. 228). There are "tendencies" in ordinary language (that is, presumably, in its ordinary speakers) to treat

- (a) "It is false that the S is P " as *logically equivalent* to "The S is not P ."
- (b) "The S is P " as the *contradictory* of "The S is not P ," and " SaP " as the contradictory of " SoP ."
- (c) " SaP " and " SeP " as *contraries*.
- (d) Such statements as neither true nor false if their referring expressions (subject terms) do not refer (are empty).

Strawson does not wish to "canonize" these tendencies into a rigid system but it is clear that his minimum aim is to show how these "primary" logical features can be consistent in a system.

In *Introduction* (pp. 176-177) and *Presupposing* (p. 231), Strawson points out that the propositions in question have these logical relations only if each has a truth value.

In *Introduction* this claim is preceded by a lengthy examination (pp. 163-174) of alternative means of interpreting the classical relations between A , E , I and O propositions and these alternative means are found to be unsatisfactory on one or another score. So that it appears that the real justification for Strawson's rules and for L_3 is the claim that they are necessary for a consistent and plausible interpretation of the classical relations. (See p. 173, for example.)

11. *Taken in a simple form the claim is false since the legislation is insufficient.*

I shall now go on to argue essentially that Strawson betrays confusion as to the nature of his own proposals and, in particular, that he has not clearly seen what is the central condition of his system. Further that not all his rules can be made to work given the most favorable interpretation of this condition.

What is required for two propositions to be contraries, contradictories, equivalents, etc., is that compounds of them be *necessarily* true or *necessarily* false. If this condition is abandoned then the relations between the propositions are material, not logical. That Strawson accepts this is clear in

³ Rather different arguments against these recommendations may be found in a paper by Douglas Odegard, "Unique Reference and Entailment," *Analysis*, vol. 23 (1963), pp. 73-79.

many places, and most forcibly stated perhaps in *Introduction* (p. 94).

Now it is quite plain that the L_s versions of A , E , I , and O propositions do not give us what is required. Some examples: the disjunction of " SaP " and " SoP " is not necessarily true in the system: the conjunction of " SeP " and " SaP " is not necessarily false; the conjunction of " SaP " and " $\neg SiP$ " is not necessarily false. The reader may easily verify that for all the classical relations there are some cases of breakdown of this kind.⁴

Related difficulties occur in the system with the logical relation of equivalence in L_s . " SiP " and " PiS " stand in allegedly different logical relations to the same proposition, for " SiP " presupposes what " PiS " entails, viz., "There are S ." In the case of contrapositives, " SaP " entails "There are P ," to which " PaS " is logically indifferent in L_s . It neither entails nor presupposes it.

In the light of these pervasive features of the system, it seems to be anything but obvious that the constants "All" and "Some" of L_s have "just the sense which they have in a vast group of statements of ordinary speech." It seems, in fact, to be simply false.

12. *A suggestion made by Strawson to remedy this is unclear and two possible interpretations of it are false.*

Strawson is aware of these features, however, and claims to have forestalled any objection which may be thought to arise from them. For he emphasizes the following restriction, R : *it is only under the condition that the propositions have a truth value that they stand in the required relations* (see *Introduction*, p. 177, and *Presupposing*, p. 231). It is not at all clear what this suggestion, R , is supposed to be. There are only three possible interpretations, so far as I can see, two of which embody mistakes and the third, which is perfectly satisfactory in giving us what is required systematically, is quite independent of Strawson's rules for "true" and "false" and inconsistent with his rules for "entails." I can only conclude that Strawson is himself confused as to what he has suggested and I shall briefly try to justify my conclusion after examining the three possible interpretations.

First, the restriction R could be interpreted as claiming that the senses of the constants "All" and

"Some" are the required ordinary senses given that condition. But it obviously cannot be maintained that the fulfilling of the condition (it is really just the existential one) actually alters the meaning of "All" and "Some" in "All S are P ," etc. I do not think it would be at all plausible to defend the system against the charge made in the last paragraph of subsection 11 by saying that the sense of a word changes under a contingent condition of that kind. It could only be maintained that the ordinary sense of "All," etc., is such that the classical relations hold only under the fulfilling of that condition, which is really what the second and third interpretations deal with.

Second, R could be interpreted as claiming that if, e.g., the propositions " SaP " and " SoP " of L_s have a truth value then they are contradictories. That is, if " SaP " and " SoP " have truth values then " SaP or SoP " is necessarily true. If we take Strawson literally (and surely we should) I think it is clear that this is what we should take him to have said in R . This, however, is a plain mistake. What can be said is this:

(a) Necessarily (If " SaP " and " SoP " have a truth value then " SaP or SoP ").

But it certainly does *not* follow from the necessary conditional together with the contingent truth of its antecedent that

(b) Necessarily (" SaP or SoP "), which is what is required to support the claim that " SaP " and " SoP " become contradictories when the condition is fulfilled. So that, on a literal interpretation, Strawson's suggestion embodies a mistake in modal logic.

13. *A third suggestion is sufficient but not necessary. Moreover, it actually requires that S entails S^* .*

It might be thought that this is uncharitable literalness. For another interpretation of R may be appealed to which would yield a quite new system, which I shall call L'_s . In L'_s we say that " SaP " and " SoP " of L_s are contradictories *relative to* the proposition "There are S ." That is, L'_s is a logical system, based on L_s but employing notions of relative necessity and possibility which are familiar from unformalized discourse. This relative modal logic has been excellently discussed by Prof. G. H. von Wright.⁵ Now, in L'_s we may say that

⁴ An earlier version of this paper, read at the Conference of the Australasian Association of Philosophy, August, 1963, contained a detailed examination of L_s so as to display these features. But so humdrum an exercise in logic should not be allowed to burden the presses.

⁵ "A New System of Modal Logic," in *Logical Studies* (London, 1957). That this interpretation ought to be considered was made clear to me by my colleague Mr. D. C. Stove.

the logical relations required do hold, in the relative modal sense, between the L'_s analogues of the A , E , I , and O propositions.

Let me develop a little the suggestion just made. To assert that " p " is necessary relative to " q " is to assert that " $\sim p.q$ " is contradictory. Here we select a certain proposition, it may be atomic or molecular, as what I shall call the modal reference proposition, and we assert that various other propositions, atomic or molecular, are possible, necessary, etc., or stand in certain logical relations with respect to the reference proposition. This is a quite distinct proposal from the mistaken one discussed in the last paragraph but one. For there is, in relative modal logic, a parallel fallacy, viz., suppose "If p then r " is necessary relative to " q " and that " p " is true but not necessary relative to " q "; then it is fallacious to conclude that " r " is necessary relative to " q ." In L'_s we have as a modal reference proposition a conjunction of existential propositions, the conjunction asserting that the subject term of any L'_s proposition of A , E , I , or O form is non-empty. Then in L'_s all the logical relations required for the square of opposition will obtain in a relative modal sense.

But if this interpretation of Strawson's suggestion is correct, then it is false that his recommendations as to truth, falsity, and entailment are necessary to achieve the fortunate result. For, first, if L'_s is a possible system so is L'_r . This is based on L_r in which A , E , I , and O propositions are given a Russellian analysis. We get L'_r by adding to L_r the very same suggestion about relative modal logic as is adopted in L'_s . And the modal reference proposition of L'_r is the very same as that for L'_s . In L'_r all the logical relations required for the classical square of opposition hold in a relative modal sense as the reader may easily verify for himself. L'_r retains the law of excluded middle.

But a second objection is stronger. For in L'_s it is certainly the case that " SaP " and " SiP " each entails "There are S ." This holds quite independently of my arguments, in thesis (2a), that each entails it in L_s . For the entailment of "There are S " by " SaP " clearly holds relative to a reference proposition containing "There are S " as a conjunct. This is flatly inconsistent with Strawson's claim about the relation between them, so we cannot interpret his suggestion simply in the light of L'_s .

An attempt might be made to avoid this by adopting yet another system L'_s . In this system L'_s is restricted so that only certain forms of

proposition of L'_s are *wffs* of the new system, viz., only A , E , I , and O forms. It will not be the case in L'_s that " SaP " entails "There are S " since the latter is not a proposition of L'_s at all. But it will still be the case that " SaP " stands in the right relation to " SoP " in L'_s as in L_s . However, we cannot answer the question "What, then, is the relation of ' SaP ' to 'There are S ' in this system?" For there is no relation at all. The question is confused and betrays a misunderstanding as to what the system L'_s is. Certainly it would be quite wrong to say that the relation is "special and odd" (see *Referring*, p. 34). This can scarcely be accepted consistently with Strawson's view that " SaP " presupposes "There are S ."

14. Possible sources of confusion.

At this point I shall be so rash as to offer a diagnosis as to what has gone wrong. It may be just one of the last mentioned confusions; that obscurity has risen because Strawson is asking for the relation between classical and existential propositions in a context in which we are concerned only with relations between classical propositions relative to their classes being non-empty, but not seeing clearly that in *that* context his question has no answer. On this diagnosis his confusion arises out of mistaking the context of his question. If this paragraph seems vague it is to be explicated by the preceding discussions of L_s , L'_s , and L''_s .

But I have more faith in the following suggestion, viz., that Strawson has confused the first and second positions which I mentioned at the beginning of the paper. Because (a) he puts the possibility that there are no S to himself as the meta-proposition that the statements have no truth value and (b) he is also confusedly inclined to think of " SaP " as a mere meaningful sentence without propositional status if there are no S , therefore he supposes that the restriction, R , which I have been trying to interpret is a "special and odd" one and is neither of the possibilities I have discussed. But, in truth, the possibility that there are no S is not essentially a meta-proposition at all, but readily statable as an object proposition. As for the possibility (b) immediately above, that is to be answered in the discussion of the second position, to which I now turn.

V. THESIS (3)

The second position, which claims that reference is necessary for a sentence to make a statement, is false,

15. *Relation of criticism of first position to that of second position.*

If Strawson were to abandon the first position (which, clearly, he holds) and retreat to the second, then all the foregoing criticisms would be blunted against the new view of presupposition. For much of the preceding argument assumes that we may always properly ask for the relation between the proposition S and the propositions S^* and $\neg S^*$ (or $\sim S^*$). But, on the second position, if S^* is false then there is, in L_s , no such proposition as S . On the second position S^* is not a necessary condition of the truth or the falsity of S , merely, but of the very existence of the proposition S . Clearly, a quite different relation of presupposition would be at issue and quite different arguments called for to meet it.

In discussing the first position I argued that, even granted the position,⁶ it is not the case that S merely presupposes rather than entails S^* . In this section I shall argue directly against the second position itself, as follows. I will try to show that what I think are the two most prominent features of Strawson's claims about sentence and statement are insufficient to support the second position, and further, I shall produce a brief, but I think strong argument for taking the position to be untenable. Consequently, I do not, in this section, consider whether or not certain relations ought to be called entailment relations or not. For I do not grant the second position at all.

16. *First objection to the second position: two grounds offered are insufficient.*

(a) *Non-assertive utterances*

In two places, *Referring* (p. 35) and *Presupposing* (pp. 221-222), Strawson discusses cases in which it seems clearly right to say that the uttering of meaningful indicative sentences is not the making of statements. These are cases of avowed recounting or inventing of fiction; story telling, play acting, and the like. The reason why it is right not to call this statement-making is that the speaker *asserts* nothing. He does not intend that we shall believe anything he says to us, and his intention not to deceive is clear. He commits himself to nothing.

Strawson says of such cases "that the words 'true' and 'false' and the word 'statement' belong together to one way or class of ways, of using

language; but telling stories is a way of using language which falls outside this class" (*Presupposing*, p. 221). It is unclear from the texts what relevance he takes this to have to the central cases or whether he intends it to have any very direct relevance at all. A way in which it might be thought relevant is by supposing that it establishes or supports or illustrates that the sentence "The king of France in 1965 is bald" cannot be used in a way which falls within the statement-making class of ways of using language.

I cannot see how it does establish, support, or illustrate that. Surely the distinctive fact about the use of language in fiction is that the sentences are not uttered assertively. That is to say, the speaker makes it clear, in some way, that it is not his intention to assert anything. Now it seems to be quite obvious that the uttered sentence's containing a referring expression which, on that occasion, lacks a reference, is not a necessary condition of the non-assertive use of language. We make up stories about real persons and tell them explicitly as fiction. It is, I think, equally obvious that the feature in question is not sufficient either. Moreover we have Strawson's own word for this; sentences containing referring expressions which lack a reference on the occasion of utterance *can* be used assertively; it is a "correct use of language for someone to assert S "; and this will be so in the case where the person mistakenly believes that the presupposed statement S^* is true" (*Presupposing*, p. 217). Surely it is also possible if the speaker knows S^* to be false, but intends to deceive. It seems, therefore, that we may dismiss fiction as irrelevant without more words.

Finally, on this question, I wish to remark that it is quite in order to consider fiction as consisting of false propositions (with a sprinkling of true ones perhaps) which are not asserted. Statements, then, are *asserted* propositions. I shall use such phrases as "identifying a proposition" when I mean that a sentence has been used, not necessarily assertively, but so that it is quite clear what it is that has been, e.g., supposed, entertained or asserted. Of course it is quite beside the point (the point of fiction) to say that novels are false; but it need be neither absurd nor mistaken to say so. Unasserted propositions do not lack truth values, after all. The antecedent of a conditional, for example, is, in

⁶ It should be clearly noted that I have *not* attacked the first position itself but only the definition of presupposition which Strawson erects upon it. The first position is not a new view at all. It goes back at least to Frege and has also been taken up by Church and Quine. I have expressed my belief that the first position is false. But none of the present arguments are relevant to establishing this.

standard cases, not asserted but is either true or false.

(b) *Context-dependent expressions*

Much of Strawson's case about sentences and statements is devoted to establishing the following thesis, which is no longer a matter of controversy: *if a significant type-sentence contains a context-dependent expression then no one proposition is identified by that sentence on all occasions of its utterance.*

In order to identify any proposition at all by means of such a sentence we must, of course, specify a particular occasion of its utterance and the context in which it occurs, on that occasion. Moreover, the utterance (occurrence) of a sentence does not identify a proposition in an important sense, unless the sentence is uttered by a speaker of the language. However it is perfectly clear that this thesis will not support, by itself, the thesis required for Strawson's second position, viz., *In some cases a significant sentence fails to identify a proposition when uttered by a speaker of the language.* The non-controversial thesis is true and the required one false just if, for any sentence, it always identifies some proposition when uttered by a speaker, but not the same proposition on all occasions or for all speakers.

17. *Second objection to the second position: the position is false for certain centrally important cases.*

At this point, defence can constitute the introduction to attack. On the widest interpretation of what constitutes the class of context-dependent expressions, an interpretation which is I think most favorable to Strawson's case so that, e.g., proper names are included, we can, I shall argue, reconcile all of Strawson's correct claims about sentences and statements with what is, essentially, Russell's position. This may be done in two ways. First it is perfectly *possible*—not perhaps desirable—to hold that the required thesis is *false* and the non-controversial one *true* in the way mentioned at the end of the previous paragraph.

To see this, let us consider what I think is Strawson's most favorable case for the required thesis. This occurs in *Referring* (p. 38), in which a speaker presenting cupped but empty hands says "This is a fine red one." I think it is neither contradictory, absurd, nor even pointless to claim that the utterer does in fact make a false statement here; and at least that he identifies propositions which are

necessary conditions of the truth of his remark, viz., that he indicates a fine red something or other, or that there is at least one fine red thing.⁷ Consequently, since necessary conditions of the truth of his remark are false, the remark itself is false. There is, I think, nothing mistaken in this, though it is, I concede, unsatisfactorily rigid—perhaps even faintly perverse. However, there are far worse faults in the catalogue of errors than these.

But a second case, both less strenuous to defend and more interesting to consider, can be argued. Let us begin by distinguishing expressions which depend on the context to provide a determinate *sense* or *content* from those which depend on the context for whether or not they have a *reference*. In the sentence under discussion there is an expression of each kind. For "one" obviously depends upon its being linked with a word, either just previously mentioned or tacitly understood by speaker and hearer (a word such as "apple" or "postage stamp") for the *sense* or *content* of the sentence uttered in the circumstances to be sufficiently determinate. It is obvious that, in the context provided by Strawson, there is no such word, and therefore the sentence fails to provide a sufficiently determinate *sense* or *content*.

Some clarification of this is required. "Sufficiently determinate" is a relative notion to be explained as follows. If the standard use of a sentence is to provide a content more determinate than is provided on some occasion of its utterance, then on that occasion, it fails to provide a sufficiently determinate content. So it is concerned with the relation which the sense of a sentence in a particular context bears to its sense in a standard context. Further, we must distinguish the *sense* and *content* from the meaning. According to Strawson's account (*Referring*, p. 35), a sentence is meaningful if and only if it is possible to use it on some occasion to make a true or false assertion. The sentence in question is obviously meaningful therefore, but it is now clear that not only may it fail referentially on some occasion of its use, it may also fail to yield a sufficiently determinate *sense*.

The conclusion I wish to draw from this is the following one: There are occasions on which, though a speaker utters a meaningful sentence, he fails to identify a sufficiently determinate proposition. Consequently hearers cannot tell, with

⁷ I have here chosen propositions which are necessary conditions of the truth, not of the L_4 falsity of the remark. But here, of course, I am arguing outside the context of Strawson's first position.

sufficient determinateness, *what he would have them believe*, and, so far as this is true, he fails to make a *determinate* statement, or fails to identify a *determinate* proposition. Since he fails to identify a proposition as determinate as the sentence in question is *properly* used to identify (the question "Fine red what?" leaps to mind) we can justify the claim that no proposition is identified and *a fortiori* no statement made in this case.

Now I believe that if we produce the trichotomy "true, false, or lacking content," the latter alternative being explained as above, we shall produce something which is both very close to what Russell seems to have meant by "true, false, or meaningless" and also proof against any of the objections and arguments used by Strawson. We may say, following Russell, that any meaningful indicative sentence uttered on some occasion by a speaker of the language either lacks sufficient content or is true or is false. From this the theory of descriptions (or something essentially equivalent) can be reached.

It is quite clear that the sentence "The king of France is bald" will not lack a content when uttered by a speaker on any ordinary occasion, though it is perhaps possible to work up occasions upon which it would lack content. It is also clear, I have argued, that the sentence can easily be uttered assertively on such occasions. Indeed its *not* being so uttered is what calls for extraordinary circumstances.

But if this is conceded then the second position seems quite clearly to be false. For the assertive uttering of the sentence may obviously cause a hearer to believe that the king of France is bald,

and the same possibility will also be open for the assertive uttering of *any* meaningful indicative sentence on an occasion on which it yields a sufficiently determinate sense. But now, not only *may* we hold that a proposition is thus identified, we *must* hold it unless we reject the following truism: *whatever may be an object of belief is a proposition*. Furthermore, unless we reject either that statements are asserted propositions or that any proposition may be asserted, we shall be forced to say that a statement must be made on any occasion by a sentence which fulfils the contextual *content* requirements and which is uttered assertively. This throws us back upon the first position.

18. Conclusion.

Finally I wish to stress again that the import of this paper is just that the current notion of presupposition is ill-founded and any trust that is put in it is misplaced. I have not attempted to argue that there is *no* loose sense of "implies" which is weaker than "entails" or that we ought never to say that people presuppose (take for granted) the truth of various propositions. I wish only to argue that no case has been made out for the *existence* of a *new* logical relation of importance for metaphysics and logic.

It has not been my aim, either, to endorse the theory of descriptions as the only possible solution to problems of reference failure. I think it is *a* solution since it is formally consistent and workable, though I do not think it is a unique or perhaps even the best solution. But discovering the best solution is not the problem which I set myself in this paper.

University of Sydney

IV. THE MIRACULOUS

R. F. HOLLAND

MOST people think of a miracle as a violation of natural law; and a good many of those who regard the miraculous in this way incline to the idea that miracles are impossible and that "science" tells us this (the more sophisticated might say that what tells us this is an unconfused *conception* of science). I shall argue that the conception of the miraculous as a violation of natural law is an inadequate conception because it is unduly restrictive, though there is also a sense in which it is not restrictive enough. To qualify for being accounted a miracle an occurrence does not have to be characterizable as a violation of natural law. However, though I do not take the conception of miracles as violations of natural law to be an adequate conception of the miraculous, I shall maintain that occurrences are conceivable in respect to which it could be said that some law or laws of nature had been violated—or it could be said equally that there was a contradiction in our experience: and if the surrounding circumstances were appropriate it would be possible for such occurrences to have a kind of human significance and hence intelligible for them to be hailed as miracles. I see no philosophical reason against this.

But consider first the following example. A child riding his toy motor-car strays on to an unguarded railway crossing near his house and a wheel of his car gets stuck down the side of one of the rails. An express train is due to pass with the signals in its favor and a curve in the track makes it impossible for the driver to stop his train in time to avoid any obstruction he might encounter on the crossing. The mother coming out of the house to look for her child sees him on the crossing and hears the train approaching. She runs forward shouting and waving. The little boy remains seated in his car looking downward, engrossed in the task of pedaling it free. The brakes of the train are applied and it comes to rest a few feet from the child. The mother thanks God for the miracle; which she never ceases to think of as such although, as she in due course learns, there was nothing supernatural about the manner in which the brakes of the train came to be applied. The driver had

fainted, for a reason that had nothing to do with the presence of the child on the line, and the brakes were applied automatically as his hand ceased to exert pressure on the control lever. He fainted on this particular afternoon because his blood pressure had risen after an exceptionally heavy lunch during which he had quarrelled with a colleague, and the change in blood pressure caused a clot of blood to be dislodged and circulate. He fainted at the time when he did on the afternoon in question because this was the time at which the coagulation in his blood stream reached the brain.

Thus the stopping of the train and the fact that it stopped when it did have a natural explanation. I do not say a *scientific* explanation, for it does not seem to me that the explanation here as a whole is of this kind (in order for something to be unsusceptible of scientific explanation it does not have to be anything so queer and grandiose as a miracle). The form of explanation in the present case, I would say, is *historical*; and the considerations that enter into it are various. They include medical factors, for instance, and had these constituted the whole extent of the matter the explanation could have been called scientific. But as it is, the medical considerations, though obviously important, are only one aspect of a complex story, alongside other considerations of a practical and social kind; and in addition there is a reference to mechanical considerations. All of these enter into the explanation of, or story behind, the stopping of the train. And just as there is an explanatory story behind the train's stopping when and where it did, so there is an explanatory story behind the presence of the child on the line at the time when, and in the place where, he was. But these two explanations or histories are independent of each other. They are about as disconnected as the history of the steam loom is from the history of the Ming dynasty. The spacio-temporal coincidence, I mean the fact that the child was on the line at the time when the train approached and the train stopped a few feet short of the place where he was, is exactly what I have just called it, a coincidence—some-

thing which a chronicle of events can merely record, like the fact that the Ming dynasty was in power at the same time as the house of Lancaster.

But unlike the coincidence between the rise of the Ming dynasty and the arrival of the dynasty of Lancaster, the coincidence of the child's presence on the line with the arrival and then the stopping of the train is impressive, significant; not because it is very unusual for trains to be halted in the way this one was, but because the life of a child was imperiled and then, against expectation, preserved. The significance of some coincidences as opposed to others arises from their relation to human needs and hopes and fears, their effects for good or ill upon our lives. So we speak of our luck (fortune, fate, etc.). And the kind of thing that, outside religion, we call luck is in religious parlance the grace of God or a miracle of God. But while the reference here is the same, the meaning is different. The meaning is different in that whatever happens by God's grace or by a miracle is something for which God is thanked or thankable, something which has been or could have been prayed for, something which can be regarded with awe and be taken as a sign or made the subject of a vow (e.g., to go on a pilgrimage), all of which can only take place against the background of a religious tradition. Whereas what happens by a stroke of luck is something in regard to which one just seizes one's opportunity or feels glad about or feels relieved about, something for which one may thank one's lucky stars. To say that one thanks one's lucky stars is simply to express one's relief or to emphasize the intensity of the relief: if it signifies anything more than this it signifies a superstition (*cf.* touching wood).

But although a coincidence can be taken religiously as a sign and called a miracle and made the subject of a vow, it cannot without confusion be taken as a sign of divine interference with the natural order. If someone protests that it is no part of the natural order that an express train should stop for a child on the line whom the driver cannot see then in *protesting* this he misses the point. What he says has been agreed to be perfectly true in the sense that there is no natural order relating the train's motion to the child which could be either preserved or interfered with. The concept of the miraculous which we have so far been considering is distinct therefore from the concept exemplified in the biblical stories of the turning of water into wine and the feeding of five thousand people on a very few loaves and fishes. Let us call the former

the contingency concept and the latter the violation concept.

To establish the contingency concept of the miraculous as a possible concept it seems to me enough to point out (1) that *pace* Spinoza, Leibniz, and others, there are genuine contingencies in the world, and (2) that certain of these contingencies can be, and are in fact, regarded religiously in the manner I have indicated. If you assent to this and still express a doubt—"But are they really miracles?"—then you must now be questioning whether people are right to react to contingencies in this way, questioning whether you ought yourself to go along with them. Why not just stick to talking of luck? When you think this you are somewhat in the position of one who watches others fall in love and as an outsider thinks it unreasonable, hyperbolic, ridiculous (surely friendship should suffice).

* * *

To turn now to the concept of the miraculous as a violation of natural law: I am aware of two arguments which, if they were correct, would show that this concept were not a possible concept. The first can be found in chapter ten of Hume's *Enquiry Concerning Human Understanding*:

Nothing is esteemed a miracle, if it ever happen in the common course of nature. It is no miracle that a man, seemingly in good health, should die on a sudden: because such a kind of death, though more unusual than any other, has yet been frequently observed to happen. But it is a miracle, that a dead man should come to life; because that has never been observed in any age or country. There must, therefore, be a uniform experience against every miraculous event, otherwise the event would not merit that appellation. And as a uniform experience amounts to a proof, there is here a direct and full *proof*, from the nature of the fact, against the existence of any miracle; nor can such a proof be destroyed, or the miracle rendered credible, but by an opposite proof, which is superior.

The plain consequence is (and it is a general maxim worthy of our attention), "That no testimony is sufficient to establish a miracle, unless the testimony be of such a kind, that its falsehood would be more miraculous, than the fact, which it endeavours to establish; and even in that case there is a mutual destruction of arguments, and the superior only gives us an assurance suitable to that degree of force, which remains, after deducting the inferior." When anyone tells me, that he saw a dead man restored to life, I immediately consider with myself, whether it be more probable, that this person should either

deceive or be deceived, or that the fact, which he relates, should really have happened. I weigh the one miracle against the other; and according to the superiority, which I discover, I pronounce my decision, and always reject the greater miracle. If the falsehood of his testimony would be more miraculous, than the event which he relates; then, and not till then, can he pretend to command my belief or opinion.

Hume's concern in the chapter from which I have just quoted is ostensibly with the problem of assessing the *testimony of others* in regard to the allegedly miraculous. This is not the same problem as that which arises for the man who has to decide whether or not he himself has witnessed a miracle. Hume gives an inadequate account of the considerations which would influence one's decision to accept or reject the insistence of another person that something has happened which one finds it extremely hard to believe could have happened. The character and temperament of the witness, the kind of person he is and the kind of understanding one has of him, the closeness or distance of one's personal relationship with him are obviously important here, whereas Hume suggests that if we give credence to some witnesses rather than others the reason must be simply that we are accustomed to find in their case a conformity between testimony and reality (§ 89). Maybe the weakness of Hume's account of the nature of our trust or lack of trust in witnesses is connected with the fact that in some way he intended his treatment of the problem of witness concerning the miraculous to have a more general application—as if he were trying to cut across the distinction between the case where we are ourselves confronted with a miracle (or something we may be inclined to call one) and the case where other people intervene, and wanting us to consider it all as fundamentally a single problem of evidence, a problem of witness in which it would make no difference whether what were doing the witnessing were a person other than oneself, or oneself in the role of a witness to oneself, or one's senses as witnesses to oneself. This anyway is the view I am going to take of his intention here.

I can imagine it being contended that, while Hume has produced a strong argument against the possibility of our ever having certitude or even very good evidence that a miracle has occurred, his thesis does not amount to an argument against the possibility of miracles as such. But I think this would be a misunderstanding. For if Hume is

right, the situation is not just that we do not happen as a matter of fact to have certitude or even good evidence for the occurrence of any miracle, but rather that *nothing can count* as good evidence: the logic of testimony precludes this. And in precluding this it must, so far as I can see, preclude equally our having *poor* evidence for the occurrence of any miracle, since a contrast between good evidence and poor evidence is necessary if there is to be sense in speaking of either. Equally it must follow that there can be no such thing as (because nothing is being allowed to count as) discovering, recognizing, becoming aware, etc., that a miracle has occurred; and if there be no such thing as finding out or being aware (etc.) that a miracle has occurred, there can be no such thing as failing to find out or failing to be aware that a miracle has occurred either; no such thing as a discovered or an undiscovered miracle . . . *en fin*, no such thing as a miracle. So Hume's argument is, after all, an argument against the very possibility of miracles. I do not think his argument is cogent either on the interpretation I have just put upon it or on the interpretation according to which it would be an argument merely against the possibility of our having good evidence for a miracle. But before giving my reason I would like first to mention the only other line of argument which I can at present envisage against the conception of the miraculous as a violation of natural law.

Consider the proposition that a criminal is a violator of the laws of the state. With this proposition in mind you will start to wonder, when someone says that a miracle is a violation of the laws of nature, if he is not confusing a law of nature with a judicial law as laid down by some legal authority. A judicial law is obviously something which can be violated. The laws of the state prescribe and their prescriptions can be flouted. But are the laws of nature in any sense prescriptions? Maybe they are in the sense that they prescribe to us what we are to expect, but since *we* formulated the laws this is really a matter of our offering prescriptions or recipes to ourselves. And we can certainly fail to act on these prescriptions. But the occurrences which the laws are about are not prescribed to: they are simply *described*. And if anything should happen of which we are inclined to say that it goes counter to a law of nature, what this must mean is that the description we have framed has been, not flouted or violated, but falsified. We have encountered something that the description does not fit and we must therefore

withdraw or modify our description. The law was wrong; we framed it wrongly: or rather what we framed has turned out not to have been a law. The relation between an occurrence and a law of nature is different then from a man's relation to a law of the state, for when the latter is deviated from we do not, save in exceptional circumstances, say that the law is wrong but rather that the man is wrong—he is a criminal. To suggest that an occurrence which has falsified a law of nature is *wrong* would be an absurdity: and it would be just as absurd to suggest that the law has been violated. Nothing can be conceived to be a violation of natural law, and if that is how the miraculous is conceived there can be no such thing as the miraculous. Laws of nature can be formulated or reformulated to cope with any eventuality, and would-be miracles are transformed automatically into natural occurrences the moment science gets on the track of them.

But there is an objection to this line of argument. If we say that a law of nature is a description, what exactly are we taking it to be a description of? A description of what has happened up to now or is actually happening now? Suppose we have a law to the effect that all unsupported bodies fall. From this I can deduce that if the pen now in my hand were unsupported it *would* fall and that when in a moment I withdraw from it the support it now has it *will* fall. But if the law were simply a description of what has happened up to now or is happening now and no more, these deductions would be impossible. So it looks as if the law must somehow describe the future as well as the past and present. "A description of the future." But what on earth is that? For until the future ceases to be the future and becomes actual there are no events for the description to describe—over and above those that either have already taken place or are at this moment taking place.

It seems that if we are to continue to maintain that a natural law is nothing but a description then we must say that the description covers not only the actual but also the possible and is every bit as much a description of the one as it is of the other. And this only amounts to a pleonastic way of saying that the law tells us, defines for us, what is and is not *possible* in regard to the behavior of unsupported bodies. At which point we might just as well drop the talk about describing altogether and admit that the law does not just describe—it stipulates: stipulates that it is impossible for an unsupported body to do anything other than fall.

Laws of nature and legal laws, though they may not resemble each other in other respects, are at least alike in this: that they both stipulate something. Moreover the stipulations which we call laws of nature are in many cases so solidly founded and knitted together with other stipulations, other laws, that they come to be something in the nature of a framework through which we look at the world and which to a considerable degree dictates our ways of describing phenomena.

Notice, however, that insofar as we resist in this way the second of the two arguments for the impossibility of the violation concept of the miraculous and insofar as we object to the suggestion that it is possible for our laws of nature to be dropped or reformulated in a sort of *ad hoc* manner to accommodate any would-be miracle, we seem to be making the first argument—the Humean argument against the miraculous—all the stronger. For if we take a law of nature to be more than a generalized description of what has happened up to now, and if at the same time we upgrade the mere probability or belief to which Hume thought we were confined here into certainty and real knowledge, then surely it must seem that our reluctance to throw overboard a whole nexus of well-established, mutually-supporting laws and theories must be so great as to justify us in rejecting out of hand, and not being prepared to assign even a degree of probability to, any testimony to an occurrence which our system of natural law decisively rules out; and surely we shall be justified in classifying as illusory any experience which purports to be the experience of such an occurrence.

The truth is that this position is not at all justified, and we should only be landed in inconsistency if we adopted it. For if it were granted that there can be no certainty in regard to the individual case, if there can be no real knowledge that a particular event has occurred in exactly the way that it has, how could our system of laws have got established in the first place?

On Hume's view, the empirical in general was synonymous with the probable. No law of nature could have more than a degree of probability, and neither for that matter could the occurrence of any particular event. This is what gave point to the idea of a balance of probabilities and hence to his thesis about the impossibility of ever establishing a miracle. But while in the one case, that of the general law, he was prepared (in the passage from which I quoted) to allow that the probability

could have the status of a proof, in the other case he was curiously reluctant to allow this.

Now if in the interest of good conceptual sense we upgrade the probability of natural laws into certainty, so as to be able to distinguish a well-established law from a more or less tenable hypothesis, it is equally in the interest of good conceptual sense that we should upgrade in a comparable fashion the probability attaching to particular events and states of affairs, so as to allow that some of these, as opposed to others, can be certain and really known to be what they are. Otherwise a distinction gets blurred which is at least as important as the distinction between a law and a hypothesis—namely the distinction between a hypothesis and a fact. The distinction between a hypothesis and a fact is for instance the distinction between my saying when I come upon an infant who is screaming and writhing and holding his ear “he’s got an abscess” and my making this statement again after looking into the ear, whether by means of an instrument or without, and actually seeing, coming upon, the abscess. Or again it is the difference between the statement “it is snowing” when made by me now as I sit here and the same statement uttered as I go outside the building into the snow and get snowed on. The second statement, unlike the first, is uttered directly in the face of the circumstance which makes it true. I can be as certain in that situation that it is snowing as I can be of anything. And if there weren’t things of this kind of which we can be certain, we wouldn’t be able to be uncertain of anything either.

If it were remarked here that our senses are capable of deceiving us, I should reply that it does not follow from this that there are not occasions when we know perfectly well that we are not being deceived. And this is one of them. I submit that nothing would persuade you—or if it would it shouldn’t—that you are not at this moment in the familiar surroundings of your university and that in what you see as you look around this room you are subject to an illusion. And if something very strange were to happen, such as one of us bursting into flame, you’d soon know it for what it was; and of course you’d expect the natural cause to be duly discovered (the smoldering pipe which set fire to the matches or whatever it might be).

But then suppose you failed to discover any cause. Or suppose that something happened which was truly bizarre, like my rising slowly and steadily three feet into the air and staying there. You could

know that this happened if it did, and probably you would laugh and presume there must be some natural explanation: a rod behind, a disguised support beneath, a thin wire above. Or could it even be done by air pressure in some way? Or by a tremendously powerful magnet on the next floor, attracting metal in my clothing? Or if not by magnetic attraction then by magnetic repulsion? I rise in the air then, and since it is no magician’s demonstration you can and do search under me, over me, and around me. But suppose you find nothing, nothing on me and nothing in the room or above, below, or around it. You cannot think it is the effect of an anti-gravity device (even if there be sense in that idea) because there just is no device. And you know that, excluding phenomena like tornadoes, it is impossible for a physical body in free air to behave thus in the absence of a special device. So does it not come to this: that if I were to rise in the air now, you could be completely certain of two incompatible things: (1) that it is impossible, and (2) that it has happened?

Now against what I have just said I envisage two objections. The first is that my rising three feet into the air in the absence of some special cause can only be held to be an impossibility by someone who is ignorant of the statistical basis of modern physics. For example, the water in a kettle comprises a vast number of atoms in motion and anything I do to the kettle, such as tilting it or heating it, will affect the movements of these atoms. But there is no way of determining what the effect will be in the case of any single atom. It is no more within the power of physicists to predict that a particular atom will change its position in such and such a way, or even at all, than it is within the power of insurance actuaries to predict that a certain man will die next week in a road accident, or die at all. However, reliable statistical statements can be made by actuaries about the life prospects of large numbers of people taken together and somewhat similarly, statistical laws are framed by physicists about the behavior of atoms in large numbers. Statistical laws are laws of probability and it gets argued that, since this is the kind of law on which the behavior of water in a heated vessel ultimately rests, there can be no *certainly* that the kettle on the hob will boil however fierce the fire, no certainty that it will boil absolutely *every* time, because there is always the probability—infinitesimally small admittedly, but still a definite probability—that enough of the constituent atoms

in their molecules will move in a way that is incompatible with its doing so. Vessels of water and rubber balls seem to be the most frequently used examples when this argument is deployed, but the suggestion has been made to me that it (or some similar argument) could be applied to the behavior of an unsupported body near the surface of the earth, in respect of which it could be maintained that there is a certain probability, albeit a very low one, in favor of the body's having its state of rest three feet above the ground.

However, it seems to me that any such argument must rest on the kind of confusion that Eddington fell into when he said, mentioning facts about atoms as the reason, that his table was not solid but consisted largely of empty space. If you add to this that your table is in a continuous vibratory motion and that the laws governing its behavior are laws of probability only you are continuing in the same vein. To make the confusion more symmetrical you might perhaps go on to say that the movements of tables in space are only predictable even with probability when tables get together in large numbers (which accounts for the existence of warehouses). Anyway my point is that, using words in their ordinary senses, it is about as certain and as much a matter of common understanding that my kettle, when put on a fierce fire, will boil or that I shall not next moment float three feet in the air as it is certain and a matter of common understanding that my desk is solid and will continue for some time to be so. The validity of my statement about the desk is not impugned by any assertion about the behavior of atoms whether singly or in the aggregate; neither is the validity of the corresponding statements about the kettle and my inability to float in the air impugned by any assertion about the statistical basis of modern science.

The second objection grants the impossibility of a body's rising three feet into the air in the absence of a special cause and grants my certitude of this. But what I can never be certain of, the objection runs, is that all the special causes and devices that accomplish this are absent. So I am entirely unjustified in asserting the outright impossibility of the phenomenon—especially when I think to do so in the very teeth of its occurrence. My saying that it is impossible could only have the force here of an ejaculation like "Struth!" *Ab esse ad posse valet consequentia*. Supposing the thing to have occurred, our response as unguessable people should be to maintain confidence in the existence of a natural

cause, to persist indefinitely in searching for one and to classify the occurrence in the meantime as an unsolved problem. So runs the second objection.

However, the idea that one cannot establish the absence of a natural cause is not to my mind the unassailable piece of logic it might seem at first glance to be. Both our common understanding and our scientific understanding include conceptions of the sort of thing that can and cannot happen, and of the sort of thing that has to take place to bring about some other sort of thing. These conceptions are presupposed to our arguing in such patterns as "*A* will do such and such unless *X*," or "If *Z* happens it can only be from this, that or the other (kind of) cause," or "If *W* cannot be done in this way or that way it cannot be done at all." An example of the first pattern is "The horse will die if it gets no food." My rising steadily three feet in the air is a subject for argument according to the second pattern. The second pattern presents the surface appearance of being more complicated than the first, but logically it is not. Let us turn our attention to the example of the first pattern.

Suppose that a horse, which has been normally born and reared, and is now deprived of all nourishment (we could be completely certain of this)—suppose that, instead of dying, this horse goes on thriving (which again is something we could be completely certain about). A series of thorough examinations reveals no abnormality in the horse's condition: its digestive system is always found to be working and to be at every moment in more or less the state it would have been in if the horse had eaten a meal an hour or two before. This is utterly inconsistent with our whole conception of the needs and capacities of horses; and because it is an impossibility in the light of our prevailing conception, my objector, in the event of its happening, would expect us to abandon the conception—as though we had to have consistency at any price. Whereas the position I advocate is that the price is too high and it would be better to be left with the inconsistency; and that in any event the prevailing conception has a logical status not altogether unlike that of a necessary truth and cannot be simply thrown away as a mistake—not when it rests on the experience of generations, not when all the other horses in the world are continuing to behave as horses have always done, and especially not when one considers the way our conception of the needs and capacities of horses interlocks with conceptions of the needs and capacities of other living things and with a

conception of the difference between animate and inanimate behavior quite generally. These conceptions form part of a common understanding that is well established and with us to stay. Any number of discoveries remains to be made by zoologists and plenty of scope exists for conceptual revision in biological theory, but it is a confusion to think it follows from this that we are less than well enough acquainted with, and might have serious misconceptions about, what is and is not possible in the behavior under familiar conditions of common objects with which we have a long history of practical dealings. Similarly with the relation between common understanding and physical discoveries, physical theories: what has been said about the self-sustaining horse seems to me applicable *mutatis mutandis* to the levitation example also. Not that my thesis about the miraculous rests on the acceptance of this particular example. The objector who thinks there is a loophole in it for natural explanation strikes me as lacking a sense of the absurd but can keep his opinion for the moment, since he will (I hope) be shown the loophole being closed in a further example with which I shall conclude.

I did not in any case mean to suggest that if I rose in the air now in the absence of any device it would be at all proper for a religious person to hail this as a miracle. Far from it. From a religious point of view it would either signify nothing at all or else be regarded as a sign of devilry; and if the phenomenon persisted I should think that a religious person might well have recourse to exorcism, if that figured among the institutions of his religion. Suppose, however, that by rising into the air I were to avoid an otherwise certain death: then it would (against a religious background) become possible to speak of a miracle, just as it would in what I called the contingency case. Or the phenomenon could be a miracle although nothing at all were achieved by it, provided I were a religiously significant figure, one of whom prophets had spoken, or at least an exceptionally holy man.

My thesis then in regard to the violation concept of the miraculous, by contrast with the contingency concept, which we have seen to be also a possible concept, is that a conflict of certainties is a necessary though not a sufficient condition of the miraculous. In other words a miracle, though it cannot only be this, must at least be something the occurrence of which can be categorized at one and the same time as empirically certain and

conceptually impossible. If it were less than conceptually impossible it would reduce merely to a very unusual occurrence such as could be treated (because of the empirical certainty) in the manner of a decisive experiment and result in a modification to the prevailing conception of natural law; while if it were less than empirically certain nothing more would be called for in regard to it than a suspension of judgment. So if there is to be a type of the miraculous other than the contingency kind it must offend against the principle *ab esse ad posse valet consequentia*. And since the violation concept of the miraculous does seem to me to be a possible concept I therefore reject that time honored logical principle.

I know that my suggestion that something could be at one and the same time empirically certain and conceptually impossible will sound to many people ridiculous. Must not the actual occurrence of something show that it *was* conceptually possible after all? And if I contend, as I do, that the fact that something has occurred might *not* necessarily show that it was conceptually possible; or to put it the other way round—if I contend, as I do, that the fact that something is conceptually impossible does not necessarily preclude its occurrence, then am I not opening the door to the instantiation of round squares, female fathers, and similar paradigms of senselessness? The answer is that the door is being opened only as far as is needed and no farther; certainly not to instantiations of the *self-contradictory*. There is more than one kind of conceptual impossibility.

Let me illustrate my meaning finally by reference to the New Testament story of the turning of water into wine. I am not assuming that this story is true, but I think that it logically could be. Hence if anyone chooses to maintain its truth as a matter of faith I see no philosophical objection to his doing so. A number of people could have been quite sure, could have had the fullest empirical certainty, that a vessel contained water at one moment and wine a moment later—good wine, as St. John says—without any device having been applied to it in the intervening time. Not that this last really needs to be added; for that any device should have existed *then* at least is inconceivable, even if it might just be argued to be a conceptual possibility now. I have in mind the very remote possibility of a liquid chemically indistinguishable from say mature claret being produced by means of atomic and molecular transformations. The device would have to be conceived as something

enormously complicated, requiring a large supply of power. Anything less thorough-going would hardly meet the case, for those who are alleged to have drunk the wine were practiced wine-bibbers, capable of detecting at once the difference between a true wine and a concocted variety in the "British Wine, Ruby Type" category. However, that water could conceivably have been turned into wine in the first century A.D. by means of a device is ruled out of court at once by common understanding; and though the verdict is supported by scientific knowledge, common understanding has no need of this support.

In the case of my previous example of a man, myself for instance, rising three feet into the air and remaining there unsupported, it was difficult to deal with the objection that we could not be certain there wasn't some special cause operating, *some* explanation even though we had searched to the utmost of our ability and had found none. And I imagined the objector trying to lay it down as axiomatic that while there is such a thing as not knowing what the cause or explanation of a phenomenon might be there can be no such thing as establishing the absence of a cause. The example of water being turned into wine is stronger, and I would think decisive, here. At one moment, let us suppose, there was water and at another moment wine, in the same vessel, although nobody had emptied out the water and poured in the wine. This is something that could conceivably have been established with certainty. What is not conceivable is that it could have been done by a device. Nor is it conceivable that there could have been a natural cause of it. For this would have had to be the natural cause of the water's becoming wine. And water's becoming wine is not the description of any conceivable natural process. It is conceptually impossible that the wine could have been got naturally from water, save in the very strained sense that moisture is needed to nourish the vines from which the grapes are taken, and this very strained sense is irrelevant here.

"But can we not still escape from the necessity to assert that one and the same thing is both empirically certain and conceptually impossible? For what has been said to be conceptually impossible is the turning of water into wine. However, when allusion is made to the alleged miracle, all the expression 'turned into' can signify is that at one moment there was water and at a moment later wine. This is what could have been empirically certain; whereas what is conceptually im-

possible is that water should have been turned into wine if one really *means* turned into. It is not conceptually impossible that at one moment water should have been found and at another moment wine in the same vessel, even though nobody had emptied out the water and poured in the wine." So someone might try to argue. But I cannot see that it does any good. To the suggestion that the thing is conceivable so long as we refrain from saying that the water *turned into* the wine I would reply: either the water turns into the wine or else it disappears and wine springs into existence in its place. But water cannot *conceivably* disappear like that without going anywhere, and wine cannot *conceivably* spring into existence from nowhere. Look at it in terms of transformation, or look at it in terms of "coming into being and passing away"—or just look at it. Whatever you do, you cannot make sense of it: on all accounts it is inconceivable. So I keep to the position that the New Testament story of the turning of water into wine is the story of something that could have been known empirically to have occurred, and it is also the story of the occurrence of something which is conceptually impossible. It has to be both in order to be the miracle-story which, whether true or false, it is.

That expression "the occurrence of something which is conceptually impossible" was used deliberately just then. And it will be objected, no doubt, that to speak of something which is conceptually impossible is to speak of a nullity. To ask for an example of something that is conceptually impossible is not (I shall be told) like asking for a sample of a substance and you cannot in order to comply with this request produce anything visible or tangible; you cannot point to an occurrence. Indeed you cannot, strictly speaking, offer a description either: you can only utter a form of words. What I have been arguing in effect is that there is a contradiction in St. John's "description" of the water-into-wine episode. But if so, then nothing has really been described; or alternatively something has been—one should not say misdescribed but rather garbled—since a conceptual impossibility is *ex vi termini* one of which sense cannot be made.

I would reply to this that sense can certainly be made of a conceptual impossibility in the respect that one can see often enough that there is a conceptual impossibility there and also, often enough, what kind of a conceptual impossibility it is and how it arises. We can see there is an inconsistency; and words, moreover, are not the

only things in which we can see inconsistency. Human actions can be pointed to here quite obviously. And I am maintaining that there is also such a thing as making sense, and failing to make sense, of *events*. If the objector holds that in the case of events, unlike the case of human actions, sense must always be there although one perhaps fails to find it, I ask: how does he know? Why the *must*? It is not part of my case that to regard a sequence of events as senseless or miraculous is to

construe it as if it were a sort of action, or to see the invisible hand of a super-person at work in it. I have contended that there are circumstances in respect to which the expression "occurrence of something which is conceptually impossible" would have a natural enough use, and I have offered three examples. I think the expression "violation of a law of nature" could also be introduced quite naturally in this connection; we could even speak of "a contradiction in our experience."

*The University College of
Swansea, Wales*

V. WHAT WE SAY

RICHARD G. HENSON

SEVERAL years ago Professor Stanley Cavell¹ defended a view as to the status of a claim made by a native speaker about how he and his fellow native speakers talk—a view which has the welcome consequence that what “ordinary language philosophers” say about their language does not require them to leave their armchairs. This view has recently been attacked by Jerry A. Fodor and Jerrold J. Katz.² I shall not attempt a full summary of Cavell’s paper or a full defense: I do not agree with all of it that I think I understand. But the arguments put forth by Fodor and Katz, while clearly and persuasively stated, seem to me to be often mistaken or inconclusive. I shall state their major arguments in the order in which they occur, numbering them consecutively throughout.

Cavell was concerned to show that Professor Benson Mates had been wrong about the methods necessary for determining “what we say”; and he was faced with a case, discussed by Mates, which saw Ryle and Austin making incompatible claims about our use of “voluntary” and “voluntarily.” Ryle had said, with certain qualifications, that we apply these words only to actions which seem to be someone’s fault.³ Austin, on the other hand, had remarked that “we may join the army or make a gift voluntarily. . . .”⁴ Cavell agrees that Austin’s examples show that Ryle was mistaken. The main point at issue, though, is the logical, or epistemic, character of certain statements which one makes about his own language, and in particular whether one needs empirical evidence for such statements. Cavell writes:

... native speakers of English . . . do not, in *general*, need evidence for what is said in the language; they are the source of such evidence. It is from them that the descriptive linguist takes the corpus of utterances on the basis of which he will construct a grammar of

that language . . . but in general, to tell what is and isn’t English, and to tell whether what is said is properly used, the native speaker can rely on his own nose; if not, there would be nothing to count. (M, pp. 174–175).

Here Fodor and Katz offer two criticisms:

(1) What Cavell misses is the distinction between what a native speaker says . . . and what he says *about* what he and other native speakers say.

... What Cavell has failed to show is precisely that the possibility of an empirical description of a natural language presupposes the truth of the metalinguistic claims of its speakers. (W, p. 60).

(2) In respect to the kind of knowledge one has of his own language, Fodor and Katz assert that there is no difference between its grammar and semantics on the one hand and its sound system on the other.

... any argument showing that the native speaker has special license to statements about the syntax and semantics would show also that he is similarly licensed to statements of the analogous form about the sound system. But this constitutes a *reductio ad absurdum* of such an argument because, *inter alia*, it entails that a native speaker of English could never be wrong (or at least could not very often be wrong) about how he pronounces (we pronounce) an English word (or spells one?). (W, p. 61).

Point (1) represents the view of Fodor and Katz on the general issue at stake in their paper and mine. Some of our metalinguistic remarks are indisputably wrong; but I postpone general discussion of which ones, and how, and what to make of it. As to (2), it seems evident that on certain questions concerning the sound system, a

¹ In “Must We Mean What We Say?” *Inquiry*, vol. 1 (1958), pp. 172–212; referred to hereinafter as ‘M.’ This was a reply to Professor Benson Mates’s “On the Verification of Statements about Ordinary Language,” in the same issue. Both papers are now reprinted in *Ordinary Language*, ed. V. C. Chappell (Englewood Cliffs, Prentice-Hall, 1964).

² “The Availability of What We Say,” *Philosophical Review*, vol. 72 (1963), pp. 57–71; referred to here as ‘W.’ Fodor and Katz are also concerned, in this article, with some of Cavell’s remarks from “The Availability of Wittgenstein’s Later Philosophy,” *Philosophical Review*, vol. 71 (1962), pp. 67–93; referred to here as ‘A.’

³ *The Concept of Mind* (New York, Barnes and Noble, 1949), p. 69.

⁴ In “A Plea for Excuses,” reprinted in his *Philosophical Papers* (Oxford, 1961), p. 139.

native speaker should prove nearly infallible and on others not.⁵ ⁶

In Cavell's paper, a good deal hinges on the similarities and differences between two kinds of statements, typified by *S*: "When we ask whether an action is voluntary we imply that the action is fishy" and *T*: "Is *X* voluntary?" implies that *X* is fishy."

... though they are true together and false together, [they] are not everywhere interchangeable; the identical state of affairs is described by both, but a person who may be entitled to say *T*, may not be entitled to say *S*. Only a native speaker of English is entitled to the statement *S*, whereas a linguist describing English may, though he is not a native speaker of English, be entitled to *T*. What entitles him to *T* is his having gathered a certain amount and kind of evidence in its favor. But the person entitled to *S* is not entitled to *that* statement for the same reason. He *needs* no evidence for it. It would be misleading to say that he *has* evidence for *S*, for that would suggest that he has done the sort of investigation the linguist has done, only less systematically, and this would make it seem that his claim to know *S* is very weakly based. And it would be equally misleading to say that he does *not* have evidence for *S*, because that would make it appear that there is something he still needs, and suggests that he is not yet entitled to *S*. But there is nothing he needs, and there is no evidence (which it makes sense, in *general*, to say) he has: the question of evidence is irrelevant. (M, p. 182; quoted on W, p. 62).

In his claim that statements *S* and *T*, though "true together and false together," are not everywhere interchangeable and not epistemically justified in the same way, Fodor and Katz claim that Cavell makes two mistakes:

(3) His first is "to suppose that, granting that *S* and *T* are true together and false together, anything whatever follows just from the fact that *S* and *T* are not everywhere interchangeable. ... No two morphemically distinct linguistic forms are everywhere interchangeable preserving all properties of context, *not even two synonymous versions of S*." (W, p. 63).

(4) His second mistake "consists of an outright contradiction," in that he (i) grants that *S* and *T*

are "true together and false together," i.e., that they are (as Fodor and Katz choose to put it) materially equivalent; (ii) says that *T* is subject to empirical confirmation and disconfirmation; (iii) says that empirical evidence is irrelevant to *S*. But from their material equivalence it follows that "any evidence which disconfirms *T ipso facto* disconfirms *S* and that any evidence which confirms *T* likewise confirms *S*." (W, p. 61).

Point (3) is an *ignoratio elenchi*. The very passage which Fodor and Katz quote shows that Cavell does not claim that the important differences between *S* and *T* can be *inferred* from the fact that *S* and *T* are not everywhere interchangeable.

What Fodor and Katz intend in (4) is partly right and partly wrong. From the fact that two statements are materially equivalent it does not in general follow that what confirms or disconfirms one does the same for the other: what disconfirms "a Russian invented the telephone" does not disconfirm "Raphael designed St. Marks Cathedral," although these are materially equivalent; and both are equivalent to "three times three is twelve," to whose truth-value *no* empirical evidence is relevant. (I was so bewitched by Fodor and Katz on first reading that I needed my colleague, David W. Bennett, to point out this feature of confirmation and material equivalence to me.)

But this is perhaps unfair to Fodor and Katz: their argument is marred by the fact that they choose to represent Cavell's description of propositions like *S* and *T* as "true together and false together" in terms of material equivalence, while in fact neither Cavell nor they were concerned with propositions which are connected in so weak a fashion. The letters "*S*" and "*T*" are not variables in this discussion, but names—although they represent, informally speaking, any pair of an indefinitely large class of pairs of expressions any of which could have served in the discussion just as well as *S* and *T*. It is presumably this latter fact which tempts Cavell to speak of them as true together and false together—for strictly speaking, *S* and *T* must each be either true or false, not swinging hand in hand from truth to falsity and back. Very well: let us grant that Fodor and

⁵ I am unable to assess the evidence (which, according to footnote 10 of W, p. 61, is contained in M. Halle's "Phonology in a Generative Grammar," *Word*, vol. 18 (1962), pp. 54-72) for the entailment claimed in the above quotation from Fodor and Katz. Assuming that they are right in this claim, I suggest that a distinction between aspects of our knowledge of the sound system analogous to the distinction I shall draw below in regard to our semantical and syntactical knowledge will meet their argument.

⁶ It is not clear to me that "native speaker" is exactly the characterization which is needed here; having nothing better to offer, I follow Cavell. Surely *some* who are not (genetically speaking) native speakers would do as well; but of course generalizations about native speakers are not falsified if the same things can be said about some who are *not* native speakers.

Katz are thinking of pairs of expressions such that, within each pair, the members are not just materially equivalent, but are related as *S* and *T* are: i.e., are so related that they not only do, but must, have the same truth-value. Granted that they may not have intended to lay down a general principle about confirmation and mutual entailment—granted, that is, that they were not obliged to consider anything except the propositions *S* and *T*—I suggest that if Cavell is indeed guilty of a contradiction, then other pairs of expressions related as *S* and *T* are would generate a contradiction if the same things were said about them as he says about *S* and *T*; and I want to dispel the illusion that (4) is decisive by sketching a nearly parallel case in which a similar argument will be seen to fail. I apologize for the excessive familiarity of what I shall say about the parallel case.

Let *U* be my utterance "My tooth aches." Let *V* be someone else's utterance (on the same occasion) "Henson's tooth aches." I take it that *U* and *V* are "true together and false together"—i.e., that depending on the occasion of utterance, *U* will be sometimes true and sometimes false, and that *V* will always have the same truth-value as *U*. Fodor might have some evidence in favor of *V*; it would *ipso facto* be evidence (for Fodor) that I was telling the truth in saying *U*. But it would be evidence *for Fodor*, and for Katz and for any other similarly situated observer; I am *not* similarly situated. I submit that the entire passage which I have quoted above (from *M*, p. 182) would be perfectly correct if Cavell were speaking of *U* and *V* instead of *S* and *T* (and if other concomitant variations were introduced—reading "person with a toothache" for "native speaker," "dentist" for "linguist," etc.).

So much should help to dispel the illusion that (4) is decisive; it remains to show (i) that Cavell was right in what he said about *S* and *T*, not just that I am right about what could be said on similar lines about *U* and *V*; (ii) *how* the persuasive schema of (4) is mistaken. The former task must be attempted *ambulando*. The latter task then: the crucial weakness of that schema is in its omission of the fact that what confirms or disconfirms a given proposition depends in part upon the situation of the person to whom the "evidence" is presented. In saying this I am not confusing confirmation with success in getting someone to believe something. What is (for you) good evidence that

I have a toothache is not just as a matter of fact irrelevant to the strength of my belief, because I have already made up my mind or have enough evidence without it—I have *no* evidence, rather than too much of it to pay any more attention, and I am logically debarred from treating that "evidence" as evidence *for me*. Similarly, in some cases I know something simply because I remember doing it, or seeing it, and what is genuine evidence for you that it happened is entirely irrelevant for me. Indeed, memory is fallible: *sometimes* I need evidence to corroborate what I (at least seem to) remember; but sometimes I do not. (See discussion of point (6) below.)

Now Fodor and Katz are right, in part: if *T* is proven to be false, and if *S* entails *T*, *S* is false; if *T* is proven to be true, and if *T* entails *S*, *S* is true. Similar things could be said of *V* and *U*; but it does not follow that evidence confirming *V* is evidence *for me* that my utterance, *U*, is true or probably so. Thus also for *T* and *S*.

So much, I am confident, is consistent with Cavell's view—but it includes a weighty concession to his critics. Neither Cavell nor they seem to me to be entirely right. When a native speaker says something like *S*, he does not normally say it on the basis of empirical evidence, much or little; but what if something like *T* should (at least seem to) be disconfirmed by empirical evidence? (I will henceforth use "*S_n*" and "*T_n*" as variables, representing any pair of statements related as *S* and *T* are.) There are several cases to consider:

(i) *S_n* (and thus *T_n*) may in fact be true, even though some evidence is uncovered which seems to count against them. This case is worth noticing only because it is worth remembering that disconfirmation is generally inconclusive.

(ii) The speaker may be ignorant of features of dialects other than his own, so that what he says about what "we" say will be incompatible in those respects with *T_n* unless *T_n* is restricted to that speaker's dialect. This case does not, I think, vitiate Cavell's argument: I know of no case in which differences of dialect have led to an *S*-like claim's being mistaken in a philosophically interesting way. *The Concept of Mind* is not weakened by Ryle's failure to notice such expressions as "... the pore gentleman was mental. ... For an 'ole hour, 'e went on something chronic." (A sample of "ordinary language" cited by Bertrand Russell.)⁷ In the unlikely case that differences between dialects of the same language turn out

⁷ In "The Cult of Ordinary Usage," *British Journal for the Philosophy of Science*, vol. 3 (1953), p. 305.

not to be inter-translatable, and to have some conceptual significance, this would be of great interest: but even this would not tend to show that information from a user of one of those linguistic-conceptual systems about his own system is mistaken.

(iii) It may happen that T_n is decisively shown to be false by the evidence which a linguist gathers. Could such evidence induce the native speaker to withdraw his S_n ? Well, what is to count here as empirical evidence? When one does withdraw such a claim, it is often because someone has offered him a case of a certain description and asked him whether he (or "we" or "one") would be willing to ask or say so-and-so about it. Suppose, for instance, that I have noticed that to say "Jones is a capable fellow" is to say something good of him, to say that he can generally be expected to succeed at things at which he, and probably other people, want him to succeed. And suppose I say "We don't say that something is capable of something, unless we regard that something as good. We say '*liable* to err,' but '*capable* of success'." My generalization is of course false. Someone might get me to see that it is false by asking "What about 'I think he's capable of murder'?" Presented with this suggestion, I might very well modify my initial claim. Now have I been presented with *evidence* against that claim? I have not necessarily, on this occasion, (a) observed someone using (as distinct from raising a question about the use of) the expression, (b) been given any research reports on its use, (c) looked it up in Fowler, (d) remembered myself or another using it.⁸ Presented with the example, I have realized that that is a perfectly proper use of the word, such as I or any other native speaker might employ. (See M, p. 174, second paragraph.) An outsider might have as evidence against my initial generalization either the fact that native speakers do use the phrase "capable of murder" with some frequency, or the fact that they report to him that they would be quite willing to use it, that it does not sound odd, etc. But I did not weigh any such evidence in coming to realize that I had been mistaken: I did not count my own nose.

A harder case remains. (iv) Suppose I have uttered S (or some sentence S_n) and am presented with statistical data against T (or T_n). Well, what shape is this evidence supposed to be in? Does it take the form "On such-and-such occasions, native

speakers were found to discuss the question whether certain actions were voluntary but did not consider the action in question in any way fishy"? If so, I might be quite unmoved by the data, and sensibly so: I might wonder how the investigator was sure the actions were not considered fishy, who (the investigator?) introduced the word "voluntary," and so on. Or do the data take the form of more detailed specification of the circumstances of utterance of "voluntary"? In this case, they might serve me exactly as well as—and in the same way as—the examples that someone might present for my consideration in connection with the previous case, e.g., the phrase "capable of murder." Suppose, though, that the data gained by other people from close observation of the speech of my language community do in fact conflict with some statement S_n , and suppose that the specific counter-examples are described to me in full—suppose then that I do not budge, that I still claim that they are improper or do not make sense.

Now the fact that this is logically possible does not establish that it happens; and a more careful treatment of this whole problem would include detailed analyses of several kinds of cases in which it seems to or does happen. But I reluctantly concede that it probably does. If so, this is a serious weakness in my position; the most serious, I hope, because it seems very serious indeed. I offer for consideration two relevant facts and an argument; I shall discuss the matter further in connection with point (7) below. One fact is that some people simply speak more carefully than others, and some have a keener ear for the linguistic proprieties. I do not see the significance of this fact clearly enough to know whom it helps; but it is a fact. Another fact is that when philosophers are discussing and especially when they are arguing about the use of an expression, they are likely to be in a peculiarly bad condition for getting it right, and this for at least two reasons: they let their philosophical prejudices get in the way, and they suffer from sheer excess of concentration. Compare the distortion of one's normal perceptions which may occur when one stares at a familiar object—or repeats a familiar word over and over—until it becomes strange. Perhaps our trouble here comes from the fact that in such cases, as in arguments about what we say, the usual background fades away, with resultant distortion in that which is at

⁸ Note the considerable difference between "remembering someone's saying 'capable of murder'," and "remembering that one can say 'capable of murder'."

the center of attention.⁹ But I confess peculiar dissatisfaction with what I have been saying. Does it reduce to the assertion that when such a conflict between native speakers occurs, it is because either (a) they lack sufficient erudition and/or are not paying close enough attention to what they are saying or (b) they are too sophisticated and/or are paying too close attention? (I do not have a handy index of optimum sophistication and/or closeness of attention.)

The argument bearing on the point (as promised in the preceding paragraph) is this: in order to *understand* and thus in order to evaluate evidence against any statement S_n which I might make, I should of course have to rely on my knowledge of the language in which it is presented to me; and that language might be my own and could not be a language which I know better than "my own." (This is tautological.) If S_n concerned a locution with which I am sufficiently familiar, then, evidence which counted against my S_n might strike so deeply at my confidence in my knowledge of my own language that it would be simply impossible for me to accept: it would count as strongly against my understanding the sentences in which it (the evidence) was formulated as against my original claim.¹⁰

Cavell had said that it would be extraordinary if we were often wrong in statements about what we say, made in the first person plural present indicative; "they are sensibly questioned only where there is some special reason for supposing what I say about what I (we) say to be wrong; only here is the request for evidence competent." (M, p. 183). In this connection, Fodor and Katz accuse Cavell of several mistakes.

(5) He says that type 2 statements¹¹ can be sensibly questioned only where we have special reason to think them false, whereas Fodor and Katz point out that "we often question statements, and sometimes demand evidence for them, because we know of no reason why they should be true." (W, p. 65).

(6) They add:

If we are only usually right, then we are sometimes wrong. But, then, it is *always* competent to request evidence to show that *this* is not one of those times. Whether in any particular case a statement is in fact questioned and evidence demanded is a matter of the positive utility of being right and the negative utility of being wrong. (W, p. 65).

(What Fodor and Katz presumably mean here is "Whether in any particular case a statement ought to be questioned. . . .")

(7) He holds that we are not often wrong in what we say about our own language; Fodor and Katz grant this for type 1 statements but not for type 2 statements, since a type 2 statement is "a kind of theory . . . an abstract representation of the contextual features which determine whether a word is appropriately used." (W, p. 65). Fodor and Katz claim that even the literature of the ordinary language philosophers is rich in disagreements on type 2 statements—witness Ryle and Austin on "voluntary." They remark that Cavell does not discuss cases in which there is a flat disagreement on such a statement; and they claim that such disagreements cannot be resolved by reference to the relevant type 1 assertions, "since the same kind of conflict can arise there too." (W, p. 66).

Point (7) is the one which has great weight; I shall deal first with (5) and (6), which I think have little. As to (5), when I make a type 1 (or even a type 2) statement, I can hardly be in the position of having no reason to think that it is true: I am a component of the "we" whose practice I am reporting, and I have learned the practice from other members of that group. Point (6) seems weak, even outside the realm of the special kind of knowledge under discussion. I am sometimes wrong in my computations in elementary arithmetic; does it follow that I may now be wrong in saying three and four are seven, or that it is "competent" to require special investigation in this case? I am sometimes mistaken in matching

⁹ Thus the strenuous efforts of Wittgenstein, Malcolm, Austin, and other analysts to minimize this danger by constant reminders of the circumstances in which we do use certain philosophically troublesome locutions.

¹⁰ I am indebted to Robert C. Coburn for suggesting this consideration; whether my way of developing it is in harmony with his intentions is of course another question.

¹¹ Cavell had remarked that philosophers make three types of statements about ordinary language:

(1) There are statements which produce *instances* of what is said in a language ("We do say . . . but we don't say —"; "We ask whether . . . but we do not ask whether —"); (2) Sometimes these instances are accompanied by *explications*—statements which make explicit what is implied when we say what statements of the first type instance us as saying ("When we say . . . we imply [suggest, say] —"; "We don't say . . . unless we mean —"). Such statements are checked by reference to statements of the first type. (M, 173).

I omit the third type because it seems to play no role in the subsequent arguments.

a given person with a given name; does it follow that I can now sensibly raise the question whether my own name is Richard, or my daughter's Elizabeth? My general objection to Fodor and Katz here may be put as follows: from the fact that one is sometimes mistaken in his assertions about members of a class of entities K , it does not follow that he may on just any occasion be mistaken in just any assertion about some member k of that class. For k may belong to a more or less clearly delimited subclass of K , concerning which subclass he is never mistaken, or never mistaken without there being, on the occasion of his mistake, special features which prompt doubt. (In delirium or amnesia, perhaps I *would* be mistaken about or ignorant of my daughter's name.)

The seventh point is the crucial one. Some consideration of the claim that a type 2 statement is a "theory" will help us see an important difference which Cavell seems to me to slight, though he is clearly not oblivious to it. The difference in question is between (i) generalizing as to what we mean when we say so-and-so, and (ii) recognizing cases *described in some detail* as ones in which, if one said so-and-so, he would be taken to mean or not to mean such-and-such. Utterances of both these kinds are generalizations; the former are more general, in that they abstract from the more or less detailed sketch of circumstances which distinguishes the latter. Since an account of circumstances can be indefinitely detailed, there is perhaps no difference in principle between the two kinds of utterances; but there is more difference in their epistemic status than one might expect from the difference in generality. The former kind is indeed, as Fodor and Katz claim, "a kind of theory," and of course one might over-generalize about it. The latter, if theory at all, is theory of a very special kind.

Suppose someone asks me whether, when I count, I say "five" before "six," or vice versa. If I take the question seriously at all, I suppose I could resolve it right there by counting past six, and finding which I did. Am I then *theorizing* about my own practice? Well, I don't say to myself "First I say 'one'; then I say 'two' . . ."—I just say "One, two . . ." This makes for an enormous difference between the status of my report on how I count and the status of my report, say, on how I tie my shoelaces. The latter action is perfectly familiar, but to my fingers rather than to my tongue. If I had to describe it, I should do it either through an effort of imagination or by actually

tying them and reporting my actions step by step; but either way, the action described would be different from the act of describing it. But if asked how I count, I do not (need not) perform two different actions at all.

I shall turn in a moment to some of the relevant respects in which counting is an atypical use of language. But what I have tried to bring out is that in some circumstances, we tell (or show) how we would perform some linguistic act simply by doing it. And when we are presented with a suitably detailed sketch of a situation and asked what we would say, or what we would imply, suggest, or whatnot, in saying it, we are very close to that kind of case. Thus the vast difference between *this* sort of question about how we talk, and questions about what is *generally* meant by the use of a given expression. It is a familiar fact that we can generalize after candid and careful reflection and be wrong, even when the materials which should have shown us that we were wrong are in some sense accessible to us, (sometimes) in memory. "I don't believe I have ever known of a star football player who was also an outstanding middle-distance runner." "But how about Ollie Matson?" Of course: once reminded, I realize that I know, and in some sense have known all along, that Matson was both a star football player and an outstanding quarter-miler, thus an outstanding middle-distance runner. No research was needed to persuade me that my remark was wrong, no expert testimony; simply a reminder. I do not mean to suggest that this item of knowledge, temporarily inaccessible to me, was not empirical; I am trying rather to bring out the fact that we sometimes generalize incorrectly even when the knowledge we need to show our mistake is all but immediately available—available, so to speak, for the asking. It is abundantly clear that the same sort of thing happens when we make type 2 statements. The moral is that the proper cure for such mistakes in type 2 assertions is through "assembling reminders" consisting of detailed accounts of cases. (See footnote 9.)

We are hardly through with the question what kind of theory a type 2 statement is, however. Even in the case of counting, one might claim that I must make several empirical assumptions before I can get much good out of it. For the question was not just "Will you on this occasion say 'five' before 'six'?" but "*Do* you (regularly) say. . . ." Am I not then *assuming* that what I did on that occasion was typical of my general practice? And

is it not an empirical question whether my practice accords with that of my fellow native counters?

But how seriously can such questions be taken? Sometimes they can be seriously raised—with a child who has not yet quite mastered counting, with a person suspected to be suffering from aphasia. . . . But a community in which people counted idiosyncratically—in which each man, in counting, “had his little ways”—would be a community in which *counting* did not take place. (Can you play chess without the *moves*?). It is an empirical question whether we do have such a practice, but this is not the issue. It is an empirical question whether, on a given occasion, I am in some pathological state which prevents me from counting properly. But we are not here concerned with pathological counters; nor are philosophers of ordinary language concerned with pathological speakers.

Counting, though, is a use of language which is peculiarly favorable to my views. It lacks borderline cases, eccentricities which are only perhaps errors, etc. It is almost unique in that there is, at each step in the process, exactly one right number and so (except for minor elasticities, as between “one hundred and one” and “one hundred one”) exactly one right word or phrase to use. A similar situation prevails in respect to certain religious and legal formulae; but for the most part, there will be several different ways of saying whatever one wants to say in a given situation; and occasionally there may be no standard way of saying it. But still, though the rules are vastly more complex and flexible, these other uses of language are governed by rules, and the meaningfulness of an expression *consists in* its conformity to those rules.

The familiar “game” analogy will be useful for exposition. One who plays a game which is moderately complex and highly organized must know the rules and a good deal of its strategy and tactics. There may be *some* rules with which he is unacquainted, although of course such ignorance tends to put him at a disadvantage. There are also likely to be occasional situations not covered by the rules, such that only ambiguous (if any) guidance to player and official can be gained from the rules. But (in the moderately complex and highly organized games to which these remarks are limited) it must be very rare that a player is ignorant how or whether a rule applies to a given situation. Otherwise, he would simply be unable to play the game, and his incapacities would quickly become evident to the other players. One

who plays the game often and fairly well knows and can say how “we” play it and does not need to take surveys or (except in especially out-of-the-way cases) consult the rule-book. What a beginner, or any outsider, learns as he comes to the game is empirical in character; the rules might have been different, and of course there are official bodies which make certain changes in the rules of many games. But we are talking about an experienced player of the game and the epistemic status of his reports on how it is played.

The analogy to our uses of language seems to me to be close, differing however in at least these respects: (i) The rules of a (complex, highly organized) game are likely to be more strict and inflexible and comprehensive than the rules of language, deliberately designed to cover any situation the rule-makers can envision. Innovation is often possible in language without prior notice, so to speak, but in the rules of games it bespeaks bullying or chicanery. In this regard, the game analogy perhaps makes my case look better than it deserves; but in the following two respects, the analogy makes my case look weaker than it deserves. For the rules of most of our “language games” differ from those of other games also in that (ii) the “rule-books” for language are not used in the same way as those for games; in particular, a dictionary or a grammar is to be tested by its fidelity to the practice of the speakers, instead of violations being authoritatively established as such by the fact that a player has gone against the rule-book, as in a game. And (iii) many of the rules of games impose what might be called “external impediments” to a player’s achievement of his goals, which, of course, are also normally specified in the rule-book. (Given that the object of a football player is, at a given moment, to score a touchdown, and granting that some of the rules specify what is to count as a touchdown, there are other rules which prohibit certain kinds of blocking by his teammates, prohibit him from throwing a second forward pass in a single play, prohibit him from “hurdling” a defender, and so on. I call these “external” impediments, because they could be changed without reconstituting the game, without affecting its basic objectives or strategy. And of course such rules are changed from time to time, often to maintain an interesting balance between attack and defense in the game. But generally speaking the rules of language are not of this sort; one does not normally try to achieve certain things in the use of language, feeling its rules as impedi-

ments; it would not often make sense to *wish that the rules were different*, or that one could suspend them, so that one could achieve one's ends more readily—as it often would in a game. It is exactly through fidelity to these rules that one does achieve what he does, in most uses of language; it is through the common understanding of the rules that it becomes clear to one's listener what moves in the *language game* are being made.)

These differences—(ii) and (iii)—between the role of rules in such games as football and in our use of language, seem to me to tell strongly in favor of my position. From (ii), we see more clearly that the “native player” of a language-game is normally one of the collective arbiters of correctness, superior to any rule-book; from (iii), we can perhaps see more clearly that in the mastery of a language, one's knowledge of the common rules is not necessary merely for engaging in the activity properly, or elegantly, or efficiently, but for his engaging in it—making and appreciating its moves—at all. (Much of what I have said would be mistaken if one interpreted, e.g., frightening or amusing people as “uses of language” in this contest. These and many other things can sometimes be done without following any linguistic rules at all. See—of course—Austin's *How To Do Things With Words* (Cambridge, Mass., 1962).

Fodor and Katz next offer a battery of criticisms of what Cavell says about a case in which there appears to be a significant difference between the ways in which a pair of native speakers speak. Cavell asks us to suppose that we become convinced that someone (a baker) uses the words “inadvertently” and “automatically” interchangeably; he claims that this does not prevent someone else (a professor) from saying that when “we” use the one word, we mean something different from what we mean when we use the other; and the professor is entitled also to say to the baker “the distinction is there, in the language (as implements are there to be had), and you just impoverish what you can say by neglecting it. And there is something you aren't noticing about the world.” (M, p. 200).

Fodor and Katz point out that Cavell assumes without argument that the case is one in which the baker's use is idiosyncratic, i.e., that we already know that there is a difference in meaning between this pair of words. Cavell's discussion of it cannot be expected to throw light, then, on the difficult questions we have just been discussing, concerning the possibility of conflict between native speakers,

or between the reports of one of them and the results of empirical study of their practice. But they offer a doubtful argument against Cavell's claim that the baker's speech must be impoverished.

(8) It may be the case that English contains expressions exactly synonymous with “automatic” and “inadvertent”—indeed, they claim that “there are” such expressions “which can be constructed in English.” (W, p. 68). (I find this claim puzzling: *are* there such expressions, or is it only that they can be “constructed”?) But if there are such expressions, they say, the baker may of course use them to mark the distinction(s) which others mark by using “inadvertent” and “automatic”; so his speech is not necessarily impoverished.

Well, one who misses a distinction between such a pair of words *may* notice, nevertheless, the difference(s) which they mark; and he *might* be able to “construct” some other expression to mark the difference in question. (A Bushman who lacks words for numbers over five *might* be able to tell that five groups of four dingoes total fewer dingoes than three groups of seven.) But at what stage of discourse will the baker's rectification prove effective? When he speaks to other people? How easily will they find out (a) that he does not distinguish between the meanings of the two words, and (b) which of them he uses as the other one ought to be used (supposing, that is, that he does use *one* of them correctly and the other one like it)? And—a far harder question—what about the impoverishment of his understanding of what others say? He will not “translate” what they say into his own idiom, because he does not realize that translation is necessary. The baker's insensitivity may not lead *him* into any philosophical difficulty, because he may not engage in philosophical debate; but the remarks of Fodor and Katz suggest that they may be thinking of meaning as a private mental activity. If it were that, one could mean what he chose to mean by his words, or perhaps one could simply *mean*, without bothering to use words; although he could still not simply *understand*, without listening to and discriminating the words of others. But if we are disabused of that error, it is hard to avoid the conclusion that the baker's speech and especially his comprehension of the speech of others is seriously impoverished.

As to Cavell's allegation that the baker fails to notice something about the world, Fodor and Katz offer two objections:

(9)

First, it is simply false that we have distinct non-synonymous words for each distinction we notice. . . . Hence, from the fact that we do not have distinct words to mark a distinction, nothing follows about whether or not we notice that distinction. (W, p. 68).

And (10), even if it were true that the baker is failing to notice a distinction marked by this pair of English words, this is philosophically unimportant unless we assume that English "is a philosophically privileged language with respect to the distinctions it codes." (W, p. 68). Many natural languages code distinctions which English does not, and *vice versa*; and there are innumerable differences which could be but are not coded by any natural language. If it were the case—as they have argued it is not—that a speaker necessarily misses whatever distinctions are not coded in his (perhaps partly idiosyncratic) lexicon, then Fodor and Katz point out that every speaker of any natural language would be missing not only every distinction coded by other languages and not his own, but also the innumerable differences which are not marked by *any* natural language. The accusation that the baker is not noticing something about the world is thus "completely trivialized." (W, p. 69). It does not follow from all this, according to our authors, that the baker *cannot* be making some philosophically significant mistake in his idiosyncratic use of these words.

What these criticisms do show is that one cannot establish that a philosophically significant error has been made *simply* by showing that someone has failed to draw a distinction coded in English. Moral: showing that one ought to draw a distinction is not something that can be done just by appealing to the way speakers in fact talk. This takes doing philosophy.

This mistake of inferring "ought" statements about distinctions from "is" statements about what speakers say deserves the name "the natural language fallacy." The general philosophical importance of this fallacy is this: once the natural language fallacy has been recognized, it becomes necessary to raise seriously the question of the utility of appealing to what we ordinarily say as a means of resolving philosophical disagreements. (W, pp. 69–70).

Fodor and Katz are of course right that we notice many differences not "coded" by pairs of words in our language, and that many other languages code some of these and miss some of the differences which are coded in English. They are presumably

well aware of the fact (and this is perhaps what they allude to in their remark about "constructing" expressions) that the lack of a single word in one natural language which is exactly synonymous with a single word of another by no means establishes that the speakers of the former cannot notice or give expression to the concept carried by that word in the latter. (See their references, note 27, W, p. 69.)

But two considerations occur to me in favor of Cavell's remark that the baker would be missing something "about the world."

(i) In reply to the charge that Cavell's argument is "trivialized," let us suppose that someone reproaches me for being a teacher when I might make substantially more money, say, in real estate; I reply "You reproach me for not making a few thousand dollars more a year? How trivial that would be in comparison with the billions of dollars I should still not be making—not to mention the francs, piastres, pesos, and the untapped wealth which no one is yet making." I acknowledge that the analogy is not quite fair and the reply partly (perhaps forty per cent) facetious. Consider then any scientist who abandons a research project because what he can hope to learn is as nothing to what he will still be ignorant of. In short, to notice something worth noticing is to do something worth doing, even though one cannot notice everything worth noticing.

(ii) What I said in connection with the "impoverishment" charge applies to the argument I have numbered (9) above. Here I add only that Fodor and Katz may be right that "from that fact that we do not have distinct words to mark a distinction, nothing follows about whether or not we notice that distinction," although I should agree with Roger Brown¹⁸ that the presence of a lexical clue in a given language (e.g., a word for a particular kind of snow) probably increases the "cognitive availability" of whatever that word characterizes (e.g., the difference between that kind of snow and other kinds). But where there is a distinct lexical clue provided by the vernacular and a native speaker fails to distinguish between the relevant expressions, it seems a plausible assumption that he is failing to notice the difference. That Bushman whose language has no numbers above five cannot be assumed to be incapable of telling the difference between a pack of eight dingoes and a pack of twenty; but if one of us (whose language contains the words for five

¹⁸ *Words and Things* (Glencoe, Free Press, 1958), pp. 235–241.

and twenty and lots of other numbers) never used different number-words for different-sized packs over five, there might be reason for suspicion.

The real importance of the issues discussed in the last few paragraphs presumably lies in their bearing on "the natural language fallacy": but our authors' enunciation of the fallacy is so brief that I am (almost) at a loss what to say about it. I take it that they might be paraphrased thus: it is a mistake to think that the distinctions which are in fact coded by a given natural language are the ones, and the only ones, which one ought to notice. (This is not the only sense which their words might be given: I hope they do not mean "What a word does mean is irrelevant to what it should (be used to) mean.") If I interpret them correctly, they are clearly right in one part of this conjoint claim: but neither Cavell nor anyone else has said that such distinctions are the *only* ones which one ought to notice. (See—if this requires any support—Cavell's first full paragraph on p. 205.) But I am at a loss to see what reason could be given for denying that one ought to notice the distinctions which *are* coded in one's own language. To say they are coded in the language is to say that they are marked in the linguistic practice of those who speak the language, not just recorded in a lexicon which may be obsolete or pedantically over-refined, or whatnot. Of course one does not on every occasion want to make every distinction for which the language offers scope (if indeed that is a coherent suggestion). But in using a natural language, we are not obliged thus to use it to the hilt, so to speak; and to use it correctly, we must mark the distinctions coded by such parts of it as we are using.

Finally, Fodor and Katz consider Cavell's remark that

such questions as "What should we say if . . . ?" or "In what circumstances would we call . . . ?", asked of someone who has mastered the language . . . is a request for the person to say something about himself, describe what he does. So the different methods [of determining how we talk] are methods for acquiring self-knowledge. . . . (A, pp. 87–88).

(11) Granting that the knowledge in question is "in *some* sense self-knowledge," Fodor and Katz remark that "this has no implications for the methods we can employ in discovering such knowledge, since the knowledge we gain in correctly describing human physiology is also in *that* sense self-knowledge." (W, p. 70). And in a most telling passage:

(12)

any facet of a speaker's use of English that is not shared by other speakers is *ipso facto* not relevant to a description of English. It is perhaps Cavell's failure to grasp this principle that has led him to suppose that some special privilege accrues to statements we make about our language in the first person plural present indicative. (W, p. 70).

In the last quotation but one, Cavell speaks of the respondent as being asked to "describe what he does," and this language seems appropriate. Speaking is something we *do*, not something which happens to us or in us; we sometimes choose our words deliberately, and we seldom say what we say unintentionally. None of us is today so innocent as to think that the concept of intentional action is easy to characterize, or that what we do intentionally is *ipso facto* easy to report accurately or even honestly; but I cannot take seriously the suggestion that it is "inaccessible" in the degree and manner in which the facts of physiology are.

What I have said above, about counting and the rules of games is the rest of my answer to point (11) and most of my answer to (12). I have admitted that our use of words is very seldom as strictly uniform as our use of the numerals in counting; but it must also be admitted that to speak a language just is to speak it, with very minor aberrations, as the other members of the linguistic community speak it. The problem of dialect is certainly important here, not only in that some speakers will be familiar with special technical vocabularies unknown to others, but also in that people at one level of education will use correctly words which people at a different level will sometimes use incorrectly. But it cannot be too heavily stressed that it is not this kind of difficulty about "how we talk" that contributes to philosophical error. If a native speaker says "When we call something 'precious', we mean . . ." it is possible that he will be unfamiliar with the sense of the word in which a drama critic describes a performance as "perhaps somewhat precious." But this is simply not the kind of problem that causes trouble for philosophers, whatever it may do for lexicographers. The disagreement over "voluntary" did not arise from this sort of ignorance; and no extravagant erudition distinguishes the users of such words as "know," "see," "good," "prove," "true," "think," "mean," "explain," or "faith." Our use of such words could not, in general, be any more idiomatic than our practice of making change or keeping score in a

game. I am looking at one side of the coin which Fodor and Katz see from the other side when they say that "any facet of a speaker's use of English that is not shared by other speakers is *ipso facto* not relevant to a description of English." Fair enough; and my side of the coin reads something like "One who moves his knight like a bishop is not accepted as a chess player." (Inscriptions on coins must be brief and unqualified; some of my qualifications and explanations are in preceding parts of the paper. A large question which I have not tried to answer except partially and negatively is: What deviations from the common linguistic practice are philosophically significant? I shall try to deal with this question in another place.)

How important is this dispute about the epistemic status of our knowledge of our own language for the "ordinary language philosophy"? None of the parties to this discussion has suggested that we cannot by *any* means find out what we say, or what we mean in saying it; and is this not enough to enable the ordinary language philosophy to bear whatever burdens it must? We can answer only tentatively pending a full account of the nature

of those burdens; but I am inclined to agree with Mates, Cavell, Fodor, and Katz that the present dispute is of great importance.

Cavell has done much to bring out this importance, in passages not discussed by Fodor and Katz or by me. More should be said about this than he said or than I shall attempt here. For the present: the "oppressive" effect of the ordinary language philosophy which Cavell mentions (M, p. 172) comes partly from the fact that, with exceptions and qualifications which have been dealt with here only partially and skimpily (I have said nothing about poetic deviations, for instance), it tells us that we must mean what we say, i.e., what *is meant* by one who utters the words we utter.¹³ (I do not pause to argue this now familiar point here.) Any private intention of meaning such-and-such in using a form of words which is not accepted in the practice of the community as an appropriate bearer of that meaning is irrelevant to what is meant. But such an account of meaning can be true only if our knowledge of what we say and what we mean in saying it is—except in very special cases—immune to refutation by empirical evidence about how we talk.

University of Utah

¹³ One may say "Pardon me" and in some sense *mean* "You are very rude to stand for so long in my way"; and in other ways too numerous to deal with here, my statement requires expansion and qualification.

VI. ON BEING IN THE SAME PLACE AT THE SAME TIME

PETER CAWS

"NOBODY has ever noticed a place except at a time," says Minkowski, "or a time except at a place."¹ One might add: "and nobody has ever noticed a place except *here*, or a time except *now*." With this addition, what was meant as an innocent argument for the interdependence of space and time becomes a serious obstacle to all cosmologies in the traditional sense. This paper is an attempt to draw out some of the philosophical consequences of the fact that the observer must always be located here and now. It is a commentary on some aspects of the theory of relativity which seem still, after half a century, to be misunderstood.

The quotation from Minkowski is taken from his paper on space and time in which he introduces the *postulate of the absolute world*: "the substance at any world-point may always, with the appropriate determination of space and time, be looked upon as at rest." His decision to use the term "absolute" to describe the four-dimensional world of space-time seems curious, since the theory which led to this view of the world was a theory which promised freedom from absolutes and their replacement by relativistic determinations, but it illustrates just that ambivalence in the theory of relativity with which this paper will be concerned. The loss of absolute rest and motion in absolute space and time was a serious shock to physics, comparable to the loss (in more recent developments) of certain aspects of conservation and symmetry; in both cases the immediate reaction was to look for some way of restoring, in a slightly modified form, what had become psychologically indispensable. In the relativistic case the new version appeared to be even better than the old; the effect of Minkowski's world-postulate is, as Cassirer points out, that "the world of physics changes from a *process* in a three-dimensional world into a *being* in this four-dimensional world."² The world-postulate has its first and most obvious application at the place and

time where the observer happens to find himself, but it was of course assumed that it might be applied equally well anywhere else in the universe. The observer may be considered at rest, but then for purposes of argument any other point may just as well be considered at rest. In fact, however, considering other points as at rest is only a game—for serious scientific purposes the observer *must* be considered at rest. There is no such thing as a moving observer.

This conclusion was foreseen by as early a thinker as Nicholas of Cusa. "As it will always seem to the observer," he says, "whether he be on the earth, or on the sun or on another star, that he is the *quasi*-motionless center and that all the other things are in motion, he will certainly determine the poles of this motion in relation to himself. Thus the fabric of the world will *quasi* have its center everywhere and its circumference nowhere."³ The last sentence is as succinct a statement of the theory of relativity as could easily be found. For Cusa the quiescence of the observer poses no problem, but that is because he too believes in an absolute, namely God, in whom all opposites are reconciled—motion and rest, center and circumference, maximum and minimum. Apparent contradictions are tolerable when there is a divine guarantee of their ultimate resolution. But such mystical resources are no longer available to us, and the denial of the possibility of a moving observer—the claim that such a being is a contradiction in terms—is intended here as something more than an exemplary paradox.

The assertion appears paradoxical, in fact, only because we are all conditioned to Newtonian modes of thought. In a homogeneous three-dimensional universe all vantage points will be equivalent, and motion from one to another is possible without any distortion of phenomena. To put the same thing in another way, observers are

¹ H. Minkowski, "Space and Time," in Einstein *et al.*, *The Principle of Relativity* (London, 1923), p. 76.

² Ernst Cassirer, tr. W. C. and M. C. Swabey, *Einstein's Theory of Relativity* (bound with *Substance and Function*) (Chicago, 1923), p. 449.

³ Nicholas of Cusa, *De Docta Ignorantia*, Bk. II, ch. 12.

interchangeable. And in the Newtonian system God is over all, the generalized observer whose omnipresence is a guarantee of the universality of the laws of motion. Every Newtonian observer could take God's point of view (i.e., any point of view removed from his own) and from it regard the world, himself included, *sub specie aeternitatis*. The reference to Spinoza is deliberate; Newtonian mechanics was a physical counterpart of Spinoza's ethics, and each rested on the possibility of seeing the world in God. Unfortunately a belief in this possibility has persisted; although contemporary scientists would hardly describe it in quite that way, many of them feel that the task of science is to give an account of the world which shall be independent of any particular perspective. But this is quite impossible.

The reason why the theory of relativity was widely thought to provide another absolute account was that it did in fact offer a formulation of the laws of nature invariant between observers, whatever their state of motion with respect to one another. (It is to be remembered that it is always the *other* observer who is moving.) Laws of nature had always been thought of as rules obeyed by the universe as a whole, and an invariant formulation of them was taken to be a new and more compendious way of saying what the universe, as a whole, was like. But with the new theory came a new insight into the nature of scientific law. A law (and this is by now so familiar that it seems hardly worth repeating) is simply a generalized relationship between observations, each made at a particular time and at a particular place; and the invariance of a formulation of such a law means simply that it can be applied to sets of observations taken in different times and places with equal success. But these observations can never be mixed, and if we wish to insert data from an observation at *A'* into calculations based on observations made at *A*, they will first have to be transformed according to some set of transformation equations appropriate to the shift from *A'* to *A*. There is no law which is capable of application to the universe as a whole.

Such a law, if it existed, would in any case be far too powerful for any practical purposes. The function of laws is to provide explanations, and there is only one world which calls for explanation, namely my own world. It would be presumptuous to suppose that that constitutes more than an insignificant fragment of the world as a whole. In

my world I am always at rest. Other bodies move about, and I get information about their movements from observations made, as always, here and now; the larger their velocity with respect to me, the odder the transformations they undergo—increases in weight, the speeding up of time, the contraction of lengths, etc. I should find such changes extremely inconvenient, and it is fortunate that I am not called upon to experience them. It is not that I do not move fast enough, but that I do not move at all (for the relativistic effects of small velocities are only quantitatively different from those of large velocities, and are equally inadmissible). I may get reports from other observers who are in motion relatively to me, but I do not accept them until they have been transformed according to the equations mentioned above. Oddly enough, these reports never make any reference to the inconvenient consequences of motion from which I congratulate myself on being preserved; these appear only if I make observations from my own point of view on the physical system of the other observer regarded now not as an observer, but as an object of observation. Occasionally, it is true, other observers attribute to *me* anomalous states of motion, etc., but these attributions are contradicted by my experience and are soon corrected by applying the appropriate transformation equation.

These considerations bring up in a novel form the whole question of the relationship of theory to observation. In theory, theoretical observers may be in motion; one well known cosmological theory refers in fact to fundamental observers moving outwards from a point of mutual origin in such a way that none of them is at rest. Such theories are, however, purely hypothetical, and they have nothing to do with the real world except in so far as their consequences are projected upon the real world. To quote Minkowski again, "only the four dimensional world in space and time is given by phenomena, but . . . the projection in space and in time may still be undertaken with a certain degree of freedom."⁴ To say that the four dimensional world is *given* by phenomena is, however, to use the term "given" in a special sense, since a complex process of reasoning separates the conclusion that the world is four dimensional from the observational evidence for it. According to more familiar usage what is given by phenomena is what has to be explained, and this is done by taking a projection of a theory which is precisely *not* given by phenomena, but which is freely constructed by

⁴ Minkowski, *op. cit.*, p. 83.

the scientific imagination. For an observer at (x, y, z, t) a theory is confirmed if its projection at (x, y, z, t) agrees with his observations, i.e., if it satisfies the boundary conditions at (x, y, z, t) . It is the possibility of different projections of the same theory, according to the different space-time situations of different observers, which Minkowski asserts in the passage quoted.

But the real world can never be the world of theory; only parts of the real world may coincide more or less exactly with parts of the world of theory when the latter are submitted to boundary conditions. And this imposition of boundary conditions has to be done afresh every time an observer makes an observation. Retrospectively, the fit of theory to the real world is remarkably good, on account of the fact that (at least in principle) those elements of theory which do not fit are discarded. But every future application of theory, even of a theory which has proved itself without exception in the past, has to be validated at the time when it is made. Such validated bits and pieces of theory remain, nevertheless, the best way of grasping the real world as it presents itself to me *in* bits and pieces; for my experience, while it validates theory cognitively, validates reality existentially. "The world can not exist," says Sartre,

without a univocal orientation in relation to me. Idealism has rightly insisted on the fact that relation makes the world. But since idealism took its position on the ground of Newtonian science, it conceived this relation as a relation of reciprocity. Thus it attained only abstract concepts of pure exteriority, of action and reaction, etc., and due to this very fact it missed the world and succeeded only in making explicit the limiting concept of absolute objectivity. This concept in short amounted to that of a "desert world" or of "a world without men"; that is too a contradiction, since it is through human reality that there is a world. Thus the concept of objectivity, which aimed at replacing the in-itself of dogmatic truth by a pure relation of reciprocal agreement between representations, is self-destructive if pushed to the limit.⁵

This would still be true philosophically even if our world were really Newtonian, but in that case a self-consistent theory of *physical* objectivity would be possible, and a useful reinforcement of the philosophical point lacking.

The appeal to Sartre is again deliberate. It is not, I think, too fanciful to say that, just as Spinoza

was said to be a moral counterpart of Newton, Sartre is a moral counterpart of Einstein. Both Spinoza and Newton devised absolute deductive systems; Sartre, like Einstein, recognizes the necessity of reducing all questions to the level of the individual observer. The data of science, no less than those of ethics, require phenomenological analysis, since man in his capacity as a knower depends on his body for entry into the physical world just as basically as, in his capacity as an agent, he depends on it for entry into the moral world. As a matter of fact most intuitive objections to the thesis of the immovable observer rest on phenomenological grounds; what makes it implausible is not the theoretical possibility of motion but our frequent consciousness of it. But this "motion of the observer" always takes place with respect to a more or less confined framework, an environment which is itself taken to be at rest and which is always of modest and human dimensions. This is part of our psychological orientation to the world which we inhabit, and only goes to show that we need to feel anchored and located in a setting which is, by comparison with ourselves, stable and enduring. The change of attitude characteristic of the shift from Newtonian to relativistic science reflects a change in the answer to the question whether the comforting characteristics of this familiar and local world can be extrapolated beyond it. The belief that they can turns out historically to be tantamount to a belief in God.

In the light of contemporary science the conclusion seems inescapable that man, condemned to carry his own perspective on the world always with him—to be in the same place at the same time—is denied the vicarious view of a domesticated universe once provided by God. The trouble is that the scientists always lag behind the philosophers in their understanding of man's relation to God. While Newton clung to his conception of the "Lord over all, who on account on his dominion is wont to be called Lord God *pantokrator*, or Universal Ruler,"⁶ Spinoza had already arrived at his *Deus sive natura*; and when Sartre had come to recognize that a universal consciousness of universal being was a contradiction in terms, Einstein still held explicitly a Spinozistic view of "a superior mind that reveals itself in the world of experience."⁷ It would of course be foolish to take

⁵ Jean-Paul Sartre, tr. Hazel Barnes, *Being and Nothingness* (New York, 1956), p. 307.

⁶ Isaac Newton, General Scholium to the second edition of the *Principia*.

⁷ Albert Einstein, "On Scientific Truth," in *Essays in Science* (New York, 1934), p. 11.

this disparity of outlook too seriously, especially since some philosophers as well as scientists share Einstein's pantheistic conviction of the intelligibility of the world in an objective sense, i.e., independently of the perspective from which we view it. But if scientific theory is only the means of rendering intelligible the world as it appears to me from my irreducibly singular point of view (and any stronger claim seems to entail quite unjustifiable assumptions) then nothing is gained by

putting into theory the possibility of my own motion except a spurious and slightly megalomaniac feeling of all-inclusive understanding. Nothing is lost by it either as long as the motion is slow compared with the velocity of light; the ordinary language of local movement does not have to be given up. The foregoing argument is addressed to the relativistic case. The immobility of the observer *can* be carried through for local motion too, but for relativistic motion it *must* be.

Carnegie Corporation of New York

VII. SANTAYANA ON JAMES: A CONFLICT OF VIEWS ON PHILOSOPHY

JOHN J. FISHER*

GEORGE SANTAYANA, confident in his ability to single out the Irishmen from among the Brahmins, never saw the consummate greatness in William James of which so many of his contemporaries, rightly or wrongly, seemed to be aware.

Upon reading the charming and instructive reminiscences of James in *Character and Opinion in the United States*, *Winds of Doctrine*, and other works, one might expect the sympathetic and generous accounts to terminate with "Here was a great American philosopher." Yet no such judgment is forthcoming. Chapter three of *Character and Opinion in the United States*, as warm and genuine a memorial as a philosopher ever received, contains references to James's courtesy, his vitality and imagination, his candor, and his simple wisdom and wistful piety; yet the reader will find marks of Santayana's deep-seated revulsion to James's philosophy on every page. As a matter of fact, in one sense of the term, Santayana argues, James was not a philosopher at all.

... philosophy was not to him what it has been to so many, a consolation and sanctuary in a life which would have been unsatisfying without it. It would be incongruous, therefore, to expect of him that he should build a philosophy like an edifice to go and live in for good. Philosophy to him was rather like a maze in which he happened to find himself wandering, and what he was looking for was the way out.¹

Although there was a marked difference between Santayana's deliberate judgment and his intuitive taste with respect to Goethe and others, he always felt that with respect to William James he did not defer to current opinion at all. Santayana's published writings were not conscious efforts to perpetuate the public image of James: they were

apparently sincere attempts to portray, in as kind a light as possible, the man whom, with Whitman, he considered as representing the genuine, long silent American mind.

In this paper, I wish, primarily, to examine Santayana's appraisal of James, with particular attention to the latter's *Varieties of Religious Experience*, the marginalia in Santayana's copy being of special interest. (For permission to use these materials I am grateful to the Librarian of Houghton Library, Harvard University, and to Mr. William James, Jr.) Happily both men were articulate and outspoken, and never did the passionate criticism of the other's thought sweep away the affection that each had for the other.

In 1905 (a year in which, according to Santayana,² James's former appraisal of the younger man's philosophy had changed from "rotten" to "sound") Santayana wrote politely to James from Athens, anticipating the arrival of James and "... the pleasure of doing a little peripatetic philosophy, under your distinguished guidance, on its native heath."³ Yet, as a student, Santayana had not been greatly impressed with James as a classroom teacher. After studying Locke, Berkeley, and Hume under James, he sensed that he had learned no clear lesson from either the instructor or the authors and had difficulty in determining whether this came about because of his own immaturity or the bewilderment of James.⁴ When, now and then, pungent scraps of learning fell in golden words from the lips of the teacher, they made far less impression on Santayana than the simple intellectual honesty, the frank confessions of doubt, and the gentle kindness of the distinguished but perplexed professor. This was the James who influenced Santayana, not the later internationally renowned pragmatist. When *Pragmatism* appeared,

* Research for this paper was aided by a Temple University Faculty Research Grant.

¹ Santayana, George, *Character and Opinion in the United States* (New York, 1920), p. 92.

² "Apologia Pro Mente Sua," *The Philosophy of George Santayana*, ed. Paul Arthur Schilpp (New York, 1940), p. 538.

³ *The Letters of George Santayana*, ed. Daniel Cory (New York, 1955), p. 74.

⁴ Santayana, George, *Winds of Doctrine* (New York, 1913), p. 248.

with its unorthodox treatment of truth, Santayana was rudely shocked.

The William James who had been my master was not this William James of the later years, whose pragmatism and pure empiricism and romantic metaphysics made such a stir in the world. It was rather the puzzled but brilliant doctor, impatient of metaphysics, whom I had known in my undergraduate days, one of whose maxims was that to study the abnormal was the best way of understanding the normal.⁵

In the earliest correspondence during Santayana's years of graduate study, we note a rather fatherly James advising and often chiding the student who clearly was experiencing a series of disenchantments with philosophy. In later years James wasted no words in his evaluation. After reading *Interpretations of Poetry and Religion* he wrote to George Palmer, with instructions to forward to Santayana, a classic of outspoken criticism.⁶

What a perfection of rottenness in a philosophy. I don't think I ever knew the anti-realistic view to be propounded with so impudently superior an air. It is refreshing to see a representative of moribund Latinity rise up and administer such reproof to us barbarians in the hour of our triumph.⁷

A man unoffended by this kind of criticism would indeed be a man of little sensitivity. Yet Santayana, undoubtedly hurt by it, replied graciously and wittily on Easter, 1900: "Palmer has just sent me your delightful letter by which I see with joy that you are full of life again in this season of resurrection."⁸ In a paragraph paralleling James in frankness and lack of cavil, Santayana, confident in the authority of classical philosophy, speaks eloquently in the name of reason in his defense.

No doubt, as you say, Latinity is moribund, as Greece itself was when it transmitted to the rest of the world the seeds of its own rationalism; and for that reason there is more need of transplanting and propagating

straight thinking among the peoples who hope to be masters of the world in the immediate future.⁹

Later, after many years have gone by, we find Santayana still musing about James's comment,

James, I need hardly say, was ready enough to snap his fingers at dialectical proofs, and deeply consented to appeal to faith, and to take sporting risks in the most serious matters. . . . What he hyperbolically called perfection was, I suppose, my way of spying the self-deceptive processes, *la fonction fabulatrice*, of the inspired mind; what he called rottenness was my apparent assumption that in the direction of religion and morals, imagination was all, and there was nothing objective.¹⁰

If Santayana is correct, James appeared to be totally incapable of understanding Santayana's philosophy. On numerous occasions in the correspondence and the *Apologia Pro Mente Sua*, Santayana repeatedly suggests that James's interpretations and criticisms were impulsive and based upon almost unpredictable spurts of admiration or revulsion. Santayana's way of looking at things remained a mystery to James, whose hostility to the philosophy, along with kindness toward the person, seemed prophetic of the reception of the philosophy of Santayana in the years ahead. If James failed to understand Santayana's philosophy, however, he is hardly to be scandalized for stupidity, for in 1940 Santayana wrote,

Now I feel that the vital foundation of my philosophy has escaped almost everybody: whence a certain sense of insecurity, even in the most friendly comments.¹¹

Until his death, James maintained an attitude of antipathy toward the mysterious foundations of Santayana's philosophy. Nevertheless James always seemed to find a source of pride and satisfaction in Santayana, and spoke highly of him in all his travels. He could write to Dickinson Miller,

Santayana's book (*The Life of Reason*) is a great one, if the inclusion of opposites is a measure of greatness.

⁵ "A General Confession," *The Philosophy of George Santayana*, p. 15.

⁶ Russell, in *Portraits From Memory*, p. 94, reports that James used the "perfection of rottenness" label for Santayana's doctoral dissertation. Santayana's dissertation, accepted in 1889, was written on the philosophy of Lotze, apparently under joint supervision, James being one of the committee. Not only is it improbable that a degree would be granted in a case in which a supervisor publicly called the dissertation "rotten," but there seems to be no evidence that James ever used the expression prior to his reading *Interpretations of Poetry and Religion* in 1900. I can only assume that Russell is inaccurate in identifying the work which occasioned the remark.

⁷ *The Letters of William James*, ed. Henry James, vol. 2 (Boston, 1920), p. 122.

⁸ *Ibid.*, p. 61.

⁹ *Ibid.*, p. 62.

¹⁰ *Apologia Pro Mente Sua*, p. 499.

¹¹ *Ibid.*, p. 503.

I think it will be reckoned great by posterity. It has no *rational* foundation, being merely one man's way of viewing things. So much of experience admitted and nothing more, so much criticism and questioning admitted and nothing more. He is a paragon of Emersonianism—declare your intuitions, though no other man share them; and the integrity with which he does it is as fine as it is rare. And his naturalism, materialism, Platonism, and atheism form a rare combination of which the center of gravity is, I think, very deep. But there is something profoundly alienating in his unsympathetic tone, his "preciousness" and superciliousness.¹²

And still conclude,

But it is a great feather in our cap to harbor such an absolutely free expresser of individual convictions.¹³

The enigmatic aspects of this relationship are reduced if we think of James as Santayana considered him: a non-philosophical philosopher who approached philosophy in a primitive way—without having a philosophy—yet whose radicalisms penetrated to the roots of all human philosophies, and expressed themselves in the unfeigned thoughts and undissembled words of a generous, often too generous, soul.

As we turn to the *Varieties* and Santayana's reaction to the work, we discover that James was not the sole possessor of directness and vigor in attack. The book was, to Santayana, simply a slumming tour in the New Jerusalem.¹⁴ The chief indictment, of course, is that of James keeping his heart and mind open to the unusual and odd in religion and philosophy.

He gave a sincerely respectful hearing to sentimentalists, mystics, spiritualists, wizards, cranks, quacks and imposters—for it is hard to draw the line, and James was not willing to draw it prematurely. He thought, with his usual modesty, that any of these might have something to teach him. The lame, the halt, and the blind, and those speaking with tongues could come to him with the certainty of finding sympathy; and if they were not healed, at least they were comforted, that a famous professor should take them so seriously: and they began to feel that after all to have only one leg, or one hand, or one eye, or to have three, might be in itself no less beautiful than to have just two, like the stolid majority.¹⁵

¹² *Letters*, vol. 2, p. 234.

¹³ *Ibid.*, p. 235.

¹⁴ The same response which he experienced years later upon reading Corliss Lamont's *Immortality*. "His Jerusalem and yours," wrote Santayana to Lamont, "seem to me so very new!" *Letters*, p. 394.

¹⁵ *Winds of Doctrine*, p. 205.

¹⁶ Perry, Ralph Barton, *The Thought and Character of William James*, vol. 2 (Boston, 1935), p. 336.

There is little in the marginal comments of Santayana in his copy of the work which reflects favorably upon it. There are just two insertions which are directly favorable: one, a simple "good" opposite (p. 331) James's proposal to test saintliness by common sense, to apply a human working principle, the survival of the humanly fittest, to religious beliefs; and the other, a "Very Fine" appended to a quotation from Malwida von Meysenbug on prayer: (p. 395) "To kneel down as one that passes away, and to rise as one imperishable." The remainder of marginalia is either expository and corrective, or critical, more often than not viciously so.

A good portion of the work, of course, consists in quotations from the writings of persons purporting to have religious or quasi-religious experiences. A number of the comments of Santayana apply to these, but implicit in almost every one is an expression of regret, even horror, that a philosopher would consider these among the case histories of religious experience.

James's approach to religious experience is given early in the work, without subterfuge, and comes under Santayana's bitter attack. James was simply excessively generous toward anyone who claimed to have a religious experience. Of course, Santayana was not alone in this criticism. One of the first recorded reactions to the *Varieties* was that of President Eliot, who wrote to James two months after publication,

The only criticism which I should make upon the book bears upon some of the narrations which you assume to be statements of fact, or accounts of actual religious experience. I often found myself doubting whether these narrations answered to any facts whatever.¹⁶

James's reply in defense (August 12, 1902) was that, of course, no man's account of experience is literally true, but in a general account of religious experience, inaccuracies of detail are of no great account!

On page 356, James's discussion of Saintliness touches upon the practical necessities of resisting evil, of jails and sanctions and all the realistic measures which sound little like the saintly quality of turning the other cheek. Yet, he argues, if the world were confined to these hard-hearted,

hard-fisted methods exclusively, if there were no charity and generosity, "no one ready to be duped many a time rather than live always on suspicion . . . the world would be an infinitely worse place than it is now to live in." Two words opposite the first part of this quotation summarize Santayana's reaction: "James himself."

Later, on page 427, as James flatly asserts that the existence of mystical states absolutely overthrows the pretension of non-mystical states to dictate what we may believe, Santayana asks, "Would this argument work in favor of giving idiots and madmen a vote?"

In the first chapter, "Religion and Neurology," James makes it clear he has no intention of dismissing religious values on pathological grounds. He admits to the materialists' notion of the organic causation of religious states, but denies that this has any relevance to the spiritual significance of the state. To say that Saint Paul was an epileptic or George Fox a hereditary degenerate may be stating the truth, but to use this bit of physiological history to refute the claims of these to experiences of superior spiritual value is nonsense. Santayana's comment is,

The point, of course, is that these religious phenomena accompany *diseased* organs. The "disease" is tested by natural ideals—life, happiness, consistent knowledge, self-sustaining pleasure, etc. The criticism is, therefore, *au fond*, a legitimate one—a denunciation of mysticism from the standpoint of politics.

From the very start, therefore, Santayana believes James leaves himself unnecessarily open to duping by the stratagems of rogues, the impostures of cranks and the sincere but erroneous testimony of the mentally ill.

In the *Varieties* there are several listings of names of persons considered by James to have experienced unusually significant religious states. On page 265 he lists among the heroes Fox, Garibaldi, General Booth, John Brown, Louise Michel, and Bradlaugh. Santayana's withering scorn is expressed by "What respect James has for the contemptible." On page 376, where James states that everyone acknowledges that among the saints and spiritual heroes would be the Franciscans, the Bernards, Luthers, Loyolas, Wesleys, Channings, Moodys, Gratrays, Brookses, the Agnes Joneses, the Margaret Hallahans, the Dora Pattisons, Santayana, in silent disgust, decorates the margin with a vigorous, meaningful exclamation mark. A few lines later, recovering his speech, he comments, "None of the

miscellaneous troupe above is comparable in human worth to Alexander or to Virgil ever."

On page 249 James refers somewhat paradoxically to a man named Billy Bray as "an excellent little illiterate English evangelist," giving Santayana perfect opportunity to append, a page later opposite an account of a particularly wild holiness experience of an unnamed person, "These little Americans are illiterate as you will, and more than the 'excellent' Billy Bray." Yet it is not only the nameless and the unknown who merit the annotator's wrath. A prayer of Pascal (page 286) is scored by Santayana as "Moral atrophy following upon optimistic piety," and a paragraph from Marcus Aurelius (page 41) evokes, "Here is fatigue, moral passivity, and, in truth, immorality. And there is contradiction besides."

The heroic figures for Santayana are Pyrrho, Alexander, and Virgil. The heroes and saints for James's study are mostly in the broad Christian tradition, and Santayana is in no mood to tolerate these. "How Christianity," he writes alongside page 47, "even decayed, has annihilated virtue." The complaint against James's cases-in-point is primarily that they are little men, of small minds, enthusiastic, irrational, and, as the marginalia on page 289 so haughtily states, intellectually atrophied.

Obviously the reluctance of Santayana to admit to candidacy for religious experience all the heroes of James implies a basic difference with respect to the nature of religious experience. In fact, the ultimate philosophical differences are so great at this point that one might wonder whether either could speak meaningfully at all to the other.

When James suggests that religion must have an enthusiastic temper of espousal in regions where morality can only acquiesce (page 48), Santayana comments "Religion=optimism. No wonder that in such a view, health itself is a disease." When James discusses the relationship between religion and happiness he argues that the commonplace happinesses which we get are "reliefs" occasioned by our momentary escapes from either experienced or threatened evils. Santayana, on seeing this on page 49, adds contemptuously, "What a servile life!" When James insists (page 51) that, in the religious life, surrender and sacrifice, even unnecessary sacrifice, are positively espoused in order to increase the happiness, Santayana sees a new definition, "Religion is then enthusiasm for the impossible."

A clue to the basic difference between James and

Santayana on the nature of religion can be found on page 491. The former has been arguing that the central notion in the religious life is the interest of the individual in his private personal destiny. As Santayana sees it, "This amounts to saying that all religion is unspiritual." From James's point of view, of course, this is eminently spiritual, for "spiritual" to him merely signifies "evaluative" as opposed to "factual" (page 4). These comments were written before Santayana's celebrated works having to do with "Spirit" (*Scepticism and Animal Faith*, *The Realm of Spirit*) were published, but the germ of the doctrine that spirit is man's highest faculty directed toward eternal essences, that man lives the spiritual life when he turns from the practical problems to the detached life of contemplation, was there. Further, the third volume of *The Life of Reason* states it quite directly. Spirituality is the noble, aspiring side of religion, the devotion to the ideal. Spirituality drinks at the springs of piety, and, using this strength, looks to the future and the ideal. The poet is a far more genuine lover of spirit than the practical, egocentric moralist. Thus Santayana can note "Strange identification" on page 500 when James associates "the various feelings of the individual pinch of destiny" with "the various spiritual attitudes." Feelings of individual destiny could indeed be spiritual for James, never for Santayana.

James reminds the reader on page 38 that there must be something solemn, serious, and tender about the religious attitude. Santayana corrects: "It must be imaginative, comprehensive, profound. It must have the sobriety of large intellect, the peace of accomplished vision," and adds that Pyrrho could have made the silliness of a Renan into a sublime philosophy.

Santayana is particularly sceptical of the rather central importance given to mysticism by James, who argues (page 379) that personal religious experience has its root and center in mystical states of consciousness. He admits that these states are more like states of feeling than states of intellect (page 380), but seems to have forgotten an earlier promise that certain ". . . more intellectual phenomena may be postponed until we treat of mysticism." Santayana, with apparent delight in having caught his mentor in an awkward position, writes "If the more intellectual be mystical, what can the less intellectual be?" (page 248). On page 428 Santayana continues the same attack. James speaks of the higher mystical states pointing in directions toward which all religious sentiments

incline. Underlining *higher*, Santayana notes, "i.e., less mystical." James continues, "They tell us of the supremacy of the ideal . . ." which prompts a vigorous "Never!" from his reader.

James uses words somewhat loosely when he deals with mysticism, and his sentences do not always stand up under rigorous analysis. Santayana can only ask "Contradiction?" of James's ascription of significance, importance, and inarticulateness to mystical states (page 382). In arguing for the ineffability of mystical states, James writes (page 380) ". . . One must have been in love one's self to understand a lover's state of mind." Santayana, on the other hand, thinks "You could not *reproduce* the lover's feeling without capacity for it: but you could *understand* it, i.e., its functions, its causes and effects."

In general, Santayana seems to admit to certain phenomena corresponding to James's "dreamy states," but does not consider these (such as some of the experiences of Tennyson) as anything more than transcendental attitudes—sympathetic absorption in certain intellectual beliefs (page 384). Santayana would like James's comments on alcoholic intoxication to have a wider application. James wrote,

To the poor and unlettered, it (alcohol) stands in the place of concepts and of literature; and it is part of the deeper mystery and tragedy of life that whiffs and gleams of something that we immediately recognize as excellent should be vouchsafed to so many of us only in the fleeting earlier phases of what in its totality is so degrading a poisoning (page 387).

Santayana's simple instruction: "Apply to mysticism as a whole."

The *Varieties* marginalia repeatedly show that Santayana's evaluation of James's use of language and argument was uniformly negative. On page 41, where James rather loosely talks about the "law of the whole," Santayana asks, "Physical or moral, ideal or conventional? What slovenly concepts!" On page 55 he corrects James's use of "believing." "To 'postulate' is not to believe, but to act on a principle for which that label is the symbol." When James on page 73 speaks of intuitions coming from the deeper level of one's nature, Santayana questions, "Is the 'deep' the true or the brutish?" When James, on page 383, speaks of the mystical "eternal inner message" of the arts, the angry annotator chides, "How words can be abused. Eternal = protoplasmic." On page 447 James speaks of an omnipotent God willing

the good and securing its triumph. Santayana notes, "If he were omnipotent he would have secured it already. 'Mighty' is what James means."

Santayana much later summed up his attitude toward this aspect of James's thinking. In *The Middle Span* he wrote, "... like Bergson, he didn't trust argument where he had intuition . . . James was no draught-horse patiently pulling the scientific barge along a placid academic canal; rather a Red Indian shooting the rapids with spasmodic skill and elemental emotions."¹⁷

It comes as no surprise then to conclude that in Santayana's analysis James's weaknesses were, on the one hand, the lack of philosophic vigor, and, on the other, the generousness that extended acceptance to all, even barbarians and fools. The honest and puzzled earnestness of James's approach was a mark of philosophical waywardness to Santayana. James's insistence that the possibility of romantic surprises must be live in every serious view of the universe was, in Santayana's view, alien to the very spirit of philosophy. The openness, the detestation of boundaries, the impatience with tradition which characterized James's thinking made it evident that he was one of that crowd of moderns who had either forgotten the Greeks or never known them, and this was sufficient to justify the acrid comment that James had no notion of what the good life was.¹⁸

Generosity is an admirable personal quality. James possessed it to a remarkable degree, and Santayana admired him for it. In response to a letter from James in 1905, he wrote, "I am grateful to you for your letter: I feel how *generous* it is, and how like you."¹⁹ Yet James was extravagantly generous philosophically. In spite of the improbability of a man's doctrine, he just might be right! As Santayana reminisced,

William James was a romantic individualist, generously sympathizing with cranks, weaklings, and imposters; they were entitled to prove themselves right, if they could, and to blaze a new trail through other people's gardens. This spontaneous love of mankind overlooks the nature and fate of mankind in deference to their wishes; it overlooks the need of tradition and teamwork.²⁰

Perhaps this undoing, James's generosity, contributed to his failure to understand Santayana's

philosophy. In 1905, in a letter unfortunately not extant, James sent Santayana an evaluation of *The Life of Reason*. Santayana's reply included these revealing sentences:

You are very generous; I feel that you want to give me credit for everything good that can possibly be found in my book. But you don't yet see my philosophy, nor my temper from the inside. Your praise, like your blame, touches only the periphery, accidental aspects presented to this or that preconceived and disparate interest.²¹

Yet if James never understood the thought of Santayana, it can be argued with corresponding success that Santayana never understood the thought of James. It is doubtful that any philosophy which neglected the Greeks could make any sense at all to Santayana. He rarely mentioned Matthew Arnold, except to affirm that he had read him, but undoubtedly James was, in Santayana's eyes, a monstrous instance of Hebraism and not Hellenism. He further mistakenly associates the thoughts and hopes of James with the thinking and hoping of liberal Protestantism rather than with liberal religion of any type, a category which would not have stirred the root association of a Catholicism which Santayana never practiced and never believed, yet always seemed perversely to admire. Although Santayana's grasp of James's pragmatism might be questioned, his attitude toward it was perfectly clear. He despised it and could never tolerate a pragmatic defense of religious faith. James's philosophical democracy, in which the obscure faith healer and addict both share equal footing with the philosopher and sage, is so alien to Santayana that it is doubtful whether he ever tried to make any sense out of it.

It is not so much a question of whether or not little men have experiences which one might call religious. Santayana's criticism is essentially toward the question whether or not one ought to pay any serious attention to the aspirations and confessions of intellects which are not large or profound. He sees religion as a conscious, direct (and usually unsuccessful) pursuit of the life of reason, the seat of all ultimate values, and the testimony of history is that lofty and intense spirits have seemed to attain their highest joys in religion. The little minds, the little men are of no concern.

¹⁷ Santayana, George, *The Middle Span* (New York, 1945), p. 7.

¹⁸ *Character and Opinion in the United States*, pp. 85, 86.

¹⁹ *Letters*, p. 78.

²⁰ *Apologia Pro Mente Sua*, p. 538.

²¹ *Letters*, p. 81.

Even meager observation will show that for some men, religion is simple, uncluttered with the trappings of intellect, and religious experience results in serenity and simple joys. For others, religion is profound and imaginative, and religious experience has, in Santayana's words, "the peace of accomplished vision." If Santayana, however, is guilty of an excessively narrow view in excluding the first group, James certainly carried his generosity too far. Religion must be, for Santayana, an entirely human thing, subject to incidental mysticism as a normal disease, but never built on mysticism. Mysticism, all faith, love, and vision, has no object, and like emotion is primarily about nothing. It cannot stand alone as the basis of religion. Whereas there is no truth in religion, except as poetry, for Santayana, James sees the possibility of all sorts of truths in the religious and quasi-religious experiences of his informants.

Santayana's criticism of James's lack of discernment in allowing the alcoholic, the narcotic addict, the mystic to speak to the subject of religious experience is not only understandable but just. Yet the question concerning the relationship of religious joy to externally produced states is a legitimate one. James is quite correct in sensing that there is (infrequently, to be sure) a spiritual intoxication in religious experience akin to alcoholic intoxication. (The early Christians were accused of being drunk on the day of Pentecost.) James's mistake, as a scientist, was in failing to explore the points of difference.

The philosophical chaos which results from James's generosity is indeed distressing. If Dewey saw it, it did not weaken his judgment of James, for Dewey saw greatness in James. But Santayana saw

the chaos as the antithesis of greatness. A great man need not be right, he wrote in *Winds of Doctrine*, but he must have a firm mind, he must clarify and express what is only potential in the rest of us. He is a spontaneous variation in humanity, but not in just any direction. Mere spontaneous variation might be madness. There must be a principle of order.

How, then, should there be any great heroes, saints, artists, philosophers or legislators, in an age when nobody trusts himself, or feels any confidence in reason, in an age when the word *dogmatic* is a term of reproach? Greatness has character and severity, it is deep and sane, it is distinct and perfect. For this reason there is none of it today.²¹

Nature, Emerson once said, never sends a great man into the planet without confiding the secret to another soul. Nature apparently was not confiding about William James to George Santayana, but in other souls, less severe, less impatient, perhaps less philosophical and less brilliant, the suspicion lurked that in William James was a voice, often mistaken, misguided, and confused, yet one which should be heard. And these, rather than Santayana, were those whose criticisms and comments were most fruitful. As Royce so properly said in his *Phi Beta Kappa* oration in 1911,

You see, of course, that I do not believe James's resulting philosophy of religion to be adequate. For as it stands it is indeed chaotic. But I am sure that it can only be amended by taking it up into a larger view, and not by rejecting it. The spirit triumphs, not by destroying the chaos that James describes, but by brooding upon the face of the deep until the light comes, and with light, order. But I am sure also that we shall always have to reckon with James's view.²²

Temple University

²¹ *Winds of Doctrine*, p. 21.

²² Royce, Josiah, *William James and Other Essays on the Philosophy of Life* (New York, 1911), p. 24.

VIII. USAGE AND DUTY

A. MACC. ARMSTRONG

THE tolerant Iranian monarch Daryavahush I, according to a tale told by Herodotus (III, 38), gave certain Greeks a lesson in the power of usage. The King asked them what they would take to eat the bodies of their parents when they were dead. They replied that they would not do it for anything. He then summoned before him certain Indians, whose custom it was to dispose of their dead by eating them, and in the presence of the Greeks, who were informed of what was going on by an interpreter, he asked these Indians what they would take to burn the bodies of their parents when they were dead—cremation being the normal Greek mode of disposing of the dead. The Indians cried out in horror and implored Daryavahush not to mention such an abominable idea.

What is usage that it should have this power? It has these characteristics:

(a) It is a practice, i.e., an act performed repeatedly in a situation of a certain kind.

(b) It is a practice carried out consciously, though often unquestioningly, and not a repeated but unconscious act such as digesting food.

(c) It is a general practice, one carried out by people in general, i.e., by members of a community acting as such and not acting in a manner due to their idiosyncrasies. It is its ensemble of usages that constitutes what is common in a community and distinguishes a community from a mere number of human beings. (Of course there are communities within communities, and communities extending across communities.)

(d) It is a practice recognized as belonging to a community, i.e., a convention. It resembles an agreement in being not an unwitting but a conscious correspondence of human wills.

(e) It differs from an agreement in that whereas there have to be parties to an agreement, i.e., persons formulating some proposal and undertaking to carry it out, a convention can become established without the use of language merely by some act or practice being noticed, admired, and widely copied. The ideal on the strength of which usages become established or fall into desuetude—whether the fleeting usages called fashions or the

stable ones handed down from generation to generation—is in the last resort an ideal regarding the leading of human life or the leaving of it. This is the standard in the light of which all other standards are given importance, the standard in the light of which, for instance, the iconoclast subordinates the standard of beauty to that of holiness, and the voluptuary subordinates the standard of duty to that of pleasure.

(f) By being a practice recognized as belonging to a community, it has the force of a standard for anyone with pretensions to be a member of that community. For anyone with such pretensions it is not simply a fact but also the “done thing” as against what is “not done,” i.e., not done by the members of the community acting as such, and the doing of which stamps the doer with the stigma of being an “outsider.” A community is constituted by an ensemble of usages varying in importance in the light of the prevailing ideal of human life, so that a departure from one usage does not necessarily make the transgressor an outsider without qualification, a complete outsider.

(g) As a standard it resembles a rule in being a guide to conduct, but differs in that, as mentioned before, it does not require to be formulated. Being formulated and so universalized entails the risk at least of an incomplete specification of the kinds of situation in which the rule is meant to apply. Railwaymen and post office workers instead of going on strike for more money have sometimes maliciously “worked to rule,” slowing down their work by following the rules even in situations where the rules are not supported by usage.

There are three sorts of usage:

(a) There is linguistic usage (the Latin *consuetudo loquendi*), the generally accepted way of using words and gestures. Nobody can make words mean just what he pleases, because the meanings of words are what people in general mean by them. If, for example, anyone chose to use “square” to signify what people in general mean by “round,” he would simply not be talking English, or, to put it otherwise, he would be putting himself outside the English speech-community. Admittedly lin-

guistic usage is changeable, and a man has his own style, which consists of tiny deviations from current usage. But such deviations are only marginal, for there is a limit beyond which a style, even if still intelligible, becomes an affectation.

(b) Then there is ceremonial usage (the Latin *ritus*). Not to mention purely religious rites, there are civil ceremonies such as graduation, calling to the bar, and matrimony. The purpose of such conventional formalities and solemnities, apart from any sacramental implications, is to effectuate beyond doubt the conferring or removal of some status. The drawback, for instance, of the old status in Scotland of marriage by habit and repute was that nobody could be quite sure whether the couple were really married or only cohabiting. Ceremonial usage may become vestigial or be replaced by bare language, as in the Roman contract by stipulation, i.e., a contract clinched by the promisee asking whether the promisor promises to carry out such and such an action, and the promisor answering, "I promise." The word used for "I promise," the word required for the promise to be binding, was *spondeo*, which etymologically signifies pouring a libation, and so is evidence of a primitive ceremony clinching an agreement. But such ceremonies belong to the community, and an individual cannot, for example, fix according to his whim the manner in which a contract is entered into.

(c) Last, there is moral usage (the Latin *mos*), a customary mode of conduct recognized as belonging to a community and on that account constituting a standard for members of that community, so that a departure from the usage is a departure from standard, i.e., wrong. The disposition of a person to carry out this established mode of conduct is what is called virtue. When Plato was hunting for definitions of various virtues, he looked for the character common to standard practices. Thus in the *Laches* the first definition of bravery offered by General Laches is that it is remaining in formation and resisting the enemy without fleeing away. Under questioning, the General has to allow that this definition is too narrow, because the Scythians fight as they flee, but he explains that the Scythians are cavalymen, to whom such tactics are proper, whereas his definition applied to Greek heavily-armed infantrymen. He is then reminded that on occasion the Spartan infantrymen have taken to flight, in an attempt to break up the enemy formation (*Laches*, 190e-191c). It is only by taking for granted the virtuousness of these

practices and extracting their common feature that Plato is able to achieve, or come near achieving, a definition of bravery.

Now anyone counting himself a member of a community becomes aware of the social pressure of some usage belonging to it as soon as he is disinclined to observe it. In that case he feels the usage as an obligation, i.e., something binding him; he feels it as incumbent on him, i.e., weighing on him; he feels it as a duty, i.e., something owing from him, something that he ought to do. We should be vastly surprised if, after walking along a busy street, we were congratulated by someone on having done our duty many times, in that we had refrained from kicking or punching a full score of passers-by. We should be surprised because it had not occurred to us that we were doing our duty because we had no mind to assault passers-by. On the other hand, if ever we were so tempted, or if we were thinking of such assaults carried out by the Emperor Nero (Tacitus, *Ann.*, XIII, 25) then it would occur to us that there was such a duty. A duty is a usage, or some implication of a usage, that irks. The celebrated contrast of duty and inclination is therefore perfectly sound, though it has also to be remembered that the conflict of duty and inclination is a crisis of human life, and human life has its continuity as well as its crises.

Here the objection may be raised that a distinction must be drawn between morals and manners, only the latter being purely a matter of usage. To be sure, there is some distinction, but both manners and morals have the same essential character of disinterestedness, and the question is whether a social solecism and an abomination do not differ only in the degree of importance of the usage transgressed. There would indeed be an essential distinction and not merely one of importance, if it were possible to contrast a morality of insight with a morality of usage, but although from the time of the Greek sophists efforts have been made to conceive a morality based not on usage but on insight into the nature of things, these attempts have all failed in one of two ways: either they have discarded moral obligation along with usage, or else they have in the end fallen back upon usage.

The first attempt was that of a sophist whose theory Plato expounds in order to combat it. The sophist contrasts what is conventionally right with what is naturally right, and he urges that the former, consisting of unnatural human conventions, is worthless rubbish which restrains the strong man until he shakes it off and asserts his superior

strength over the mass of weaklings, when the light of natural right shines forth (*Gorgias*, 482c-484c; 491c-492c). On this theory the only comment that needs to be made here is that the desired shaking off of the restraints of human conventions is also the shaking off of moral obligation.

Another attempt was made by Plato, who sought to vindicate Greek conventional morality, with certain modifications, by pleading that it really was based on nature, because virtuous conduct promoted or constituted the agent's well-being, which Plato took as the ultimate aim, implanted by nature, of all human endeavor. Aristotle took much the same line. Plato's argument is that only in acting virtuously is the agent's soul in proper shape, with the rational part controlling the mettlesome and appetitive parts (*Rep.*, *passim*). Aristotle's argument is that human well-being, to be distinct from the well-being of other entities, must comprise the performance of what only man can do, or what he best can do, and this is precisely the exercise of virtue or the virtues—though Aristotle allows, against Plato, that human well-being depends partly on circumstances and can be ruined by great misfortune to the agent or his family and friends (*E.N.*, 1097a15-1101b9).

Although Plato deepens the traditional Greek conceptions of the virtues and regards goodness as something to be grasped by philosophical insight in the case of rulers and educators, his subtle definitions of bravery, sobriety, and honesty as conditions of the soul are reached by identifying with each of those conditions a character previously noticed as common to accepted practices of a certain kind. He expressly insists on the indispensability of a rigorous and testing moral training as a preliminary to philosophical insight (*Rep.*, 535a-540b).

Aristotle, again, who conceives virtue as a condition of the will situated in what is the mean, rightly calculated, between excess and defect in emotions and actions, gives as examples of virtuous action what were recognized practices. Thus he exemplifies bravery by refraining from deserting one's post or running away or throwing away one's arms, sobriety by refraining from adultery or lewdness, and gentleness by refraining from abuse

and battery (*E.N.*, 1129b20). When he elucidates bravery as a mean with regard to emotions of fear and boldness, he objects to Plato's idea that bravery is displayed not only in war but also in peril on the sea and in disease, maintaining that the term is strictly applied only in situations where there is the opportunity of showing prowess or where death is noble (*E.N.*, 1115a6-b6; 1129b20). While Aristotle declares that the landmark by which the mean is rightly calculated is fixed by moral insight, he insists that the eye of the soul does not get its power without the aid of virtue, on the ground that reasonings concerned with actions to be carried out involve a principle, and such principles are self-evident and not taught by argument, but they are not self-evident except to someone good, because wickedness twists a man and gives him delusions about them (*E.N.*, 1144a30). On Aristotle's doctrine, however, virtue is acquired by training or habituation in virtuous practices, and to be habituated is to be operative in a certain way through having been repeatedly impelled in that way by some non-instinctive guidance (*E.E.*, 1220a39; *E.N.*, 1103a25). Now if we cannot see what it is our duty to do in some situation, for instance, what amount of boldness to display, unless we have been habituated to act in that way, then our moral insight depends on the usages in which we have been habituated. The truth of this is corroborated today, when the relaxation of parental and school discipline has been followed by the occurrence of moral perplexity among the young, or in other words by the emergence of the "mixed-up kid."

Diogenes of Sinope and the Stoics took a more revolutionary line than Plato and Aristotle. They championed virtue by attempting to base it directly on nature and resolutely discounted usage, including the immemorial tabus against cannibalism, incest, etc. Diogenes regarded man's natural life as something which had been overlaid by usage,¹ insisting that life as given by the gods was easy, only it had been covered up by people's wanting honey-cakes, unguents, and the like. His test for sifting the natural from the conventional was that of economy (*euteleia*), i.e., indispensability to existence. Counting nothing important

¹ According to an account going back to his contemporary Theophrastus, Diogenes professed to have been converted to his philosophy by watching a mouse run about without looking for a bed, or being scared of the dark, or seeking any of the things reputed to be delights (Diog. Laert., VI, 22). If so, Diogenes was guilty of an error in zoology, for, as has been shown by Professor Wynne-Edwards in his *Animal Dispersion in Relation to Social Behaviour* (Edinburgh, 1962), the higher animals, especially the vertebrates and insects, mostly manage to balance their populations at or near the optimum density, so far as food resources are concerned, by adopting conventional goals, consisting of territorial rights and social status, instead of any direct struggle for food itself.

except virtue and vice, and so applying his test rigorously, Diogenes finished by renouncing not only fine clothes and such comforts as a cup, a bed, and a house, but also music, geometry, astronomy, public life, family life, politeness, and decency. Yet with this way of life virtue was emptied of content, for it was conceived not so much as a disposition to do anything as a disposition to do without, to reduce desires to the absolute minimum with the idea that then they are easy to satisfy. Diogenes' favorite comparison of his mode of life with the labors of Hercules was distinctly far-fetched, for, in fact, exercising economy of effort, he subsisted on charity in the big cities of Athens and Corinth (Diogenes Laertius, VI, 20-76; Dio Chrysostom, *Or.*, X, 29; Epictetus, *Diss.*, I, 24). The failure of his attempt to excogitate a natural morality was recognized by his contemporaries, who called him and his followers *cynics*, i.e., canines.

The Stoics, however, advocated not sallow abstinence but behaving in life as at a banquet, taking food from the dish when it is brought to you, but not reaching out for the dish before it comes or clutching it when it is being removed. According to the Stoics, nature, which is identical with God, is the reason which pervades and settles everything. Hence living in agreement with nature, which is to comply not with usage but with a universal, unchangeable law, the purely rational law of nature, entails the willing acceptance of the ineluctable decrees of Providence. Virtue, then, being the disposition to take things as they are and must be, implies not the moderation of one's emotions, as Aristotle had thought, but the absence of emotion (*apathy*), for emotion is nothing but an "irrational and unnatural movement of the soul" or, to put it otherwise, an "exaggerated urge," investing things with a glamour or repulsiveness which has to be stripped off for things to be seen as they really are. The sage who, exercising the same reason that directs the world, carries out a purely rational examination of what a thing is, what it is made of, what it is for, and how long it lasts, realizes, for instance, that Falernian wine is nothing but grape juice, pork nothing but the dead body of a pig, a purple robe nothing but sheep's wool dyed with shell-fish blood, and copulation nothing but a functional discharge. Looking at things unemotionally, the sage sees that nothing is important to him except his own virtue or vice, i.e., his success or not in treating things as they are; he is brave, honest, and sober in that he

treats no danger, no piece of property, and no bodily pleasure as being of any importance to him (Diog. Laert., VII, 87-110; Cicero, *De leg.*, I, 6, 18; Epictetus, *Ench.*, 15; M. Aurelius, *Med.*, VI, 13; XII, 18).

Yet if the sage, who satisfies the "natural" urge toward well-being by allowing nothing to ruffle the smooth flow of his existence, can watch with complete unconcern his country overrun, his property looted, and his daughters raped (Seneca, *De constantia sapientis*, 5), his supposed virtue turns out to be sheer callousness, as was remarked by the Academic philosopher Crantor. Besides, as men have sometimes to act, and so need guidance, before the decrees of Providence are patent, the Stoics had to allow that some acts, though strictly speaking unimportant, are "fitting" (*kathekonta*). On their explanation the fitting is what, when done, admits of a reasonable defence; as examples they mentioned honoring one's parents, brothers, and country, and fraternizing with friends (Diog. Laert., VII, 107-108). In this apologetic manner the Stoics began to readmit the traditional obligations, but apparently divested of any dependence on usage. For while the fitness of an action is coincident with the obligation to do it, its fitness seems to be something to which the question whether it is generally practiced is quite irrelevant.

The Stoic conception of virtue as absence of emotion derives its plausibility from a failure to distinguish between apathy, or uninterestedness, and dispassionateness, or disinterestedness. On the other hand the subsidiary Stoic notion of a natural law or right comprising acts that are fitting comes much nearer to being true. If any obligations are natural, and so universal and unchangeable, then they must be acknowledged by all civilized peoples—savages might be expected to be as benighted about obligations as about other things. If any obligations are simply fitting, then they must be apprehended not by inference but by some moral insight or appreciation. If there are some such obligations, and others can be inferred from them without recourse to usage (or statute), then the whole lot will constitute a system of natural law. Now there seems to be a handful of acts, the obligation to do which (or to refrain from doing which, refraining being to act upon oneself) not only has been acknowledged by civilized peoples for thousands of years but also is so plain that there is nothing plainer from which it could be inferred. The Ancient Iranians had as firm a grasp of the duties of keeping a promise and telling

the truth as anyone today, and the duties of refraining from theft, adultery, and murder were not only proclaimed in the Decalogue but also acknowledged by Aristotle, who in the very course of his elucidation of virtue as the mean between excess and defect points out that the acts of theft, adultery, and murder have names which directly involve badness; an adulterer, he remarks drily, is not a man who consorts with married women more than he ought (*E.E.*, 1221b18; *E.N.*, 1107a8). It is on fundamental duties like these, duties not apprehended by argument, that people rely when they proceed to argue that some act is a duty either in general or in a particular situation.

The obligation to keep a promise was examined by H. A. Prichard (*Moral Obligation*, pp. 169-179). In promising or agreeing to do some action, he observes, we appear to be creating or bringing into existence the obligation to do it, so much so that promising seems just to be making ourselves bound to do it. As, however, an obligation is a fact of a kind which it really is impossible to bring into existence directly, it must be that in what we call promising to do something what we create is not really the obligation to do the act but something else, our creation of which renders us bound to do the action. Now promising requires the actual use of the word "promise" or some equivalent such as "undertake" or "give you my word." Here it may be objected to Prichard's elucidation that promising does not require the use of "promise" or some equivalent ceremonial word; it is enough if one deliberately invites and secures the reliance of another on one's performance, particularly if one secures, as a return, the performance of a service by the other. But the making of a promise must be irrespective of its securing the reliance of the promisee on the performance of the action promised. Otherwise it would be impossible for the promisee to ask himself whether the performance of the promised action could be relied on or to think that it could not, as the Earl of Rochester wrote of King Charles II that he was one *whose word no man relies on*, because then the King could not be said even to have given his word. Nor is it true to say that inviting someone to rely on one's performance of some action is promising him to do it. Then, as Prichard remarked, an employer who on learning that his men hesitated to put their backs into their work for fear that if they did, their time rates of pay would be reduced, argued them into believing that he would not be so foolish,

would be said to be promising them not to do so. Furthermore, if he declared that he had no intention of reducing the rates, he would still not be promising them not to reduce the rates, even if his men expected that he would not change his mind, and he would be aware that he was doing something different, if, on finding them still dissatisfied, he went on to tell them that he promised not to reduce the rates if they speeded up their work.

When I promise someone to do a certain action, so Prichard continues his elucidation, I am causing him to hear a certain word, with a definite meaning for both of us, together with the term ordinarily used for the action, in such a way that he believes the sounds come from me. My doing this can give rise to an obligation to do the action, only because I have somehow or other already promised not to cause a sound of that kind in connection with the phrase for the action, without going on to do the action. However, this prior general promise or agreement cannot have been made in a way involving the use of language, for then the general agreement would itself presuppose another agreement forming the reason why the general agreement should be kept, and so on indefinitely. Prichard himself left it problematical what this mysterious something is that is implied in the existence of an agreement, which looks so much like an agreement and yet, strictly speaking, cannot be an agreement because it cannot have been brought into being by means of language, but what he unwittingly specified is a usage, in the sense elucidated previously in this article. In short, to make a promise is to go through a conventional performance referring to some future action on one's part which it is not the usage to go through without going on to do the action referred to, and so the duty to keep a promise is the duty to observe that usage.

The same is true of the other obvious duties. Stealing is taking someone's property without his consent, depriving him of it, and property just is an object which it "is done" for some person to enjoy and subject to his purposes and "not done" for anyone else to subject to his purposes without the consent of the person for whom it is "done." There does not have to be a usage whereby people own things, or, as Hegel put it, the absence of property is as little self-contradictory as property (*Phenomenology of Spirit*, V, C, c). The existence, the acquisition, and the disposal of property are all due to usage, and if there were no such institution

as property, there would be no such action as stealing. But when there is this institution constituted by usage, stealing is a departure from the usage, and the usage forms the reason why stealing is morally wrong; the wrongness, i.e., departure from standard, is inferred immediately from the departure from usage, so that the statement that it is wrong to steal has even been mistaken for a pleonasm.

The duty to refrain from murder also comes near to being a pleonasm, for murder is the wilful killing of a human being in circumstances in which it is understood that killing is wrong. What the circumstances are is left to be understood, because usage, the usage of millenia, settles that the circumstances are those other than war, judicial execution, and self-defense. It is just because the obviousness of the duty derives from its being the recognition of the usage that once the usage is challenged, e.g., in the case of judicial execution, the duty becomes a matter of controversy and loses its obviousness.

Take, again, the duty to tell the truth. There are occasions on which somebody asking a question about the situation is not owed an answer, and so the duty is, strictly speaking, the duty to refrain from misleading anyone by asserting what one considers to be false or by making an ambiguous assertion from which what is false is intended to be understood (Fichte, *Doctrine of Ethics*, V, 4, A). To mislead by making a false or ambiguous assertion is possible only because people generally make assertions on most occasions to lead someone else to believe what is the case, so far as the speaker is aware. When anyone is known to be a liar, no credence is attached to his assertions, and so he has lost the power to mislead, and if people generally were liars, no credence would be attached to anyone's assertions. So the duty in question, when its presuppositions are brought out, proves to be the duty to refrain from departing from the current practice of not asserting something to be the case, except where, so far as one is aware, it is the case.

Consider the plain duty to refrain from mating with another man's wife. This presupposes that there is such a status as being a wife. A status is not a quality, like being tall or strong or red-haired; it is a relation in which some person stands by convention to others, a relation governing mandatorily, permissively, or prohibitively that person's behavior to others and others' behavior to him.

Diogenes the Cynic, in accordance with his relentless rejection of convention, admitted in marriage nothing more than the mating of a man who makes advances with a woman who yields (Diog. Laert., VI, 72). In so far as marriage is something more than mating, a man's marrying a woman involves going through a ceremony with her which marks the assumption of a conventional status whereby none but he is to mate with her—except perhaps with his consent, as was the usage in Sparta—and so to mate with someone else's wife is to break the convention of marriage. Having gone through the ceremony the couple rank as father and mother of any offspring that they produce. To be father and mother is more than to be zoologically sire and dam; it is to possess a certain status. Thus when Plato advocated common ownership of wives and children among the members of the ruling class in his paradigmatic republic, he emphasized that what he had in mind was the extension not merely of the name of father but also of the corresponding behavior, which he expressly characterized as the respect, devotion, and obedience to parents laid down by usage (*Rep.*, 463d). Likewise the Larger Catechism applies the Fifth Commandment not only to natural parents but also to "such as, by God's ordinance, are over us in place of authority." Now to be in place of authority is to possess a status in virtue of which one is treated with respect and obedience. So the duty of honoring one's parents as such proves to be the duty of honoring those whom usage prescribes should be honored.

On examination, then, it turns out that in the case of the fundamental duties that are not taught by argument but are apprehended immediately, the very duties on which any argument about one's duty has to rely as its starting-points, the statement in full of what each duty is shows it to be the duty to observe some inveterate usage. In the light of the failure of this last attempt to conceive a morality not dependent on usage but grasped by pure insight, there is nothing for it but to abandon the sophistic supposition that the variety of usage renders it contemptible, and to adopt the view of the historian Herodotus (*loc. cit.*), who drew from the variety of usage the inference that it is foolish to despise other people's usages. The varying of usage is a consequence of its historical character, and that can seem a defect only to those who do not discern the historical character of man himself.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston	Peter Thomas Geach	Wesley C. Salmon
Alan R. Anderson	Adolf Grünbaum	George A. Schrader
Kurt Baier	Carl G. Hempel	Wilfrid Sellars
Richard B. Brandt	Jaakko Hintikka	J. J. C. Smart
Lewis W. Beck	Raymond Klibansky	Wolfgang Stegmüller
Roderick M. Chisholm	Benson Mates	Manley H. Thompson, Jr.
L. Jonathan Cohen	John A. Passmore	John Wild
James Collins	Günther Patzig	G. H. von Wright
Michael Dummett	Richard Popkin	John W. Yolton

VOLUME 2/NUMBER 2

APRIL 1965

CONTENTS

I. JEROME A. SHAFFER: <i>Recent Work on the Mind-Body Problem</i>	81	IV. G. B. KERFERD: <i>Recent Work on Presocratic Philosophy</i>	130
II. WILFRID SELLARS: <i>Meditations Leibniziennes</i>	105	V. ARTHUR C. DANTO: <i>Basic Actions</i>	141
III. HAIG KHATCHADOURIAN: <i>Vagueness, Meaning, and Absurdity</i>	119	VI. HARRY G. FRANKFURT: <i>Descartes' Validation of Reason</i>	149
		VII. GARETH B. MATTHEWS: <i>Augustine on Speaking from Memory</i>	157

UNIVERSITY OF PITTSBURGH PRESS

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

REPRINTS

Authors who are subscribers will receive 50 reprints gratis. Additional reprints can be purchased through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* for individual subscribers is six dollars and the price *per annum* for institutions is ten dollars. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues are sold at the rate of two dollars to individuals, and three dollars to institutions. Correspondence regarding subscription and back orders may be addressed directly to the publisher (University of Pittsburgh Press, University of Pittsburgh, Pittsburgh, Pennsylvania 15213).

* * *

The *American Philosophical Quarterly* is published quarterly in January, April, July, and October by the University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, Pennsylvania, 15213.
Second-class postage paid at Pittsburgh, Pennsylvania.



I. RECENT WORK ON THE MIND-BODY PROBLEM

JEROME A. SHAFFER

INTRODUCTION

THIS paper is a discussion of recent work on the mind-body problem. That problem remains a source of acute discomfort to philosophers. Few problems have received more attention, and there has been much progress, yet still no solution stands out as markedly superior. It looks like a muddle and many have proclaimed that they could prove it was. But somehow the mind-body problem has always survived, and, astonishingly enough, in much the same form. To this day we remain in doubt. What we regard as the most plausible solution at one time seems riddled with hideous difficulties the next.

The most profound, and, for that reason, most enduring solution was that of Descartes. Therefore the major progress in recent years has come from the attacks upon his doctrine. These have made more clear to us some of the places where that doctrine succeeds and where it fails. I have selected four different lines of attack for discussion. In the first section I consider Ryle's attack on the Cartesian distinction of mental and physical and what remains of the problem after that attack. In Section II, I consider the attack of the Cyberneticians on Descartes' doctrine that man cannot be regarded as an organic machine (whereas animals can be so regarded). Then in Section III, I consider the attack of Identity Theorists on Descartes' ontological dualism. Finally, in the last section, problems of causal connection between the mental and physical are taken up.

Many facets of the mind-body problem are not touched at all in this critical account. Nor are all the promising lines of approach considered. Finally, I have even had to omit discussion of some of the important recent contributions. At best this paper concerns itself with some of the ideas which have been influential in the thinking of philosophers in the last fifteen years or so.

I. THE MENTAL

Ryle's Thesis

The traditional "mind-body problem" is the

problem of the relation between the mental and the physical in humans and other higher animals. To ask about a relation presupposes two different things to be related. Many thinkers reject this presupposition and therefore the alleged "problem" which is supposed to arise from it. These thinkers do not have difficulty with the concept of the physical body; their difficulty lies in the *mental* term of the alleged relation. At the very least, for there to be a genuine problem here, it must be shown that there are expressions referring to the mental which have meaning and a different meaning from expressions referring to the physical. This is usually not shown but simply assumed to be the case. This assumption has been powerfully attacked in recent times by Gilbert Ryle. One of Ryle's main theses, expressed in an extreme form (even Ryle sometimes backs away from this extreme form, as we shall see), is as follows:

It is being maintained throughout this book that when we characterize people by mental predicates, we are not making untestable inferences to any ghostly processes occurring in streams of consciousness which we are debarred from visiting; we are describing the ways in which those people conduct parts of their predominantly public behaviour. True, we go beyond what we see them do and hear them say, but this going beyond is not a going behind, in the sense of making inferences to occult causes; it is going beyond in the sense of considering, in the first instance, the powers and propensities of which their actions are exercises.¹

Thus when we attribute some mental predicate to someone, we are attributing to him some bit of behavior (either its performance or its outcome) or a disposition toward some behavior or both. If Ryle's assimilation of the mental to behavior is legitimate, then mind and body are not different in principle and the conventional dualistic theories rest upon a confusion. Thus it is said in a recent assessment of Ryle's contribution that such theories as Interactionism and Parallelism:

are undermined simply by pointing out that mind-displaying concepts of the sort in question do not

¹ Gilbert Ryle, *The Concept of Mind* (London, 1949), p. 51 and cf. also p. 161.

allow for the supposed double stream of events at all. Hence no theory is required to explain how the two streams interact or, for that matter, fail to interact.²

Before indicating the exaggeration involved in this claim, let me say that I agree with the bulk of Ryle's program. If we confine ourselves to qualities of mind and personality and to the exercises of these qualities, to qualities of mind such as being acute or inventive, to personality features such as being ambitious or cynical, in a sulky or stubborn mood, or to the exercise of such qualities as in paying attention or obeying the law, scrutinizing or savoring, shamming or reminiscing, then Ryle is right in claiming that the essence of such qualities and their exercise lies in what people do or are prone to do, not in some double action, "one act on the physical stage and another on the mental stage."³

Sensations

It is a popular misconception about *Concept of Mind* that it attempts to analyze *all* mental predicates behavioristically. William Kneale is guilty of this misconception when he says, "Ryle finds it necessary to deny not only that there can be mental substances but also that there can be mental events, unless these latter are supposed to be pieces of bodily behavior which manifest the capacities and dispositions said to constitute minds."⁴ And such passages as the one of Ryle's quoted above support that misconception. But the truth is that Ryle himself gives a special status to two classes of mental predicates, those which refer to bodily sensations such as itches, tinglings, throbbings, aches, etc., and those which refer to feelings, of which Ryle gives as examples, "a throb of compassion," "a shock of surprise," "a thrill of anticipation," "a twinge of remorse," "a qualm of apprehension," "a pricking of conscience," "a glow of pride."⁵ These are obviously not bits of behavior. Nor are they, Ryle insists, dispositions or propensities to behave. They are genuine non-behavioral events. As Ryle says, "'I have a twinge' asserts that an episode took place."⁶ Of course, it could be either a bodily sensation like a twinge of rheumatism or a feeling like a twinge of remorse, "though the word

'twinge' is not necessarily being used in quite the same sense in the alternative contexts."⁷

In his sections on the emotions, Ryle is out to attack a prevalent view that the bulk of our emotional life consists of private occurrences. He insists on the occurrent, episodic nature of feeling and sensation to contrast them with:

inclinations and moods, including agitations, (which) are not occurrences and do not therefore take place either publicly or privately. They are propensities, not acts or states. . . . Feelings, on the other hand, are occurrences.⁸

The relegation of the bulk of our emotional life to dispositions is accomplished while admitting that feelings and sensation, at least, are not dispositions. And since they clearly are not items of behavior either, they must have some special status. Ryle does not himself draw this conclusion but it is inescapable. The predicates which refer to feelings and sensations of the sort here indicated are not analyzable into public, overt pieces of behavior, nor propensities toward such acts; therefore they must signify something private and covert. Call them (derisively) "ghostly" or "occult." If they cannot be analyzed away, still less can they be sneered away.

Thoughts

Besides sensations and feelings, what else qualifies as mental events? There is a large and heterogeneous assortment of occurrences which, following the excellent article of W. J. Ginnane,⁹ we can call *thoughts*. We report such events when we say, "As I walked in, a terrible thought occurred to me," "At midnight, the thought crossed my mind that . . .," "It suddenly came to me that . . .," "Just then it dawned on me that . . .," etc. Thoughts of this sort, as we say, "come into one's head," "go through one's mind," or sometimes "lurk at the back of one's consciousness." Such thoughts are referred to as inklings, suspicions, impressions, or realizations. Whenever we report the occurrence of a thought in one of these many ways, we are reporting the occurrence of an event which took place at some particular time. If we

² W. J. Ginnane, "Thoughts," *Mind*, vol. 69 (1960), pp. 372-390 (see p. 373).

³ Ryle, *op. cit.*, p. 63.

⁴ William Kneale, *On Having a Mind* (Cambridge, 1962), p. 34.

⁵ Ryle, *op. cit.*, pp. 83-84.

⁶ *Ibid.*, p. 209.

⁷ *Ibid.*, p. 84.

⁸ *Ibid.*, p. 83.

⁹ W. J. Ginnane, *op. cit.*

may be allowed to extend the term "thought" in the natural way that Descartes did,¹⁰ we may also add the following as reports of mental events: Suddenly I was assailed by the doubt that . . . , comprehended that . . . ; as he finished all at once I got the idea that . . . ; when I saw what they had done, I made a vow that . . . , made a resolution to . . . , came to the decision that. . . . In addition to Descartes' list of thoughts we might appropriately add memories, hopes, wishes, desires, fears, the events which the following expressions often are used to report: "Suddenly I remembered that, recalled that, recollected that . . . ," "Suddenly, I had the hope that, the wish that, the desire that, the fear that, the apprehension that. . . ." I shall say that all of these expressions can be used to report *thoughts*.

As in the case of reports of feelings, reports of thoughts are not translatable into reports of behavior or tendencies toward behavior. So far as their meaning is concerned, they are reports of genuine occurrences but not of overt performances. Ryle attempts to remove the mentalistic bite from our concept of thought by analogizing thoughts with speaking. He is in the tradition of those Behaviorists who took thinking to be "sub-vocal speech" or "laryngeal behavior" (J. B. Watson) when he describes thinking in such terms as "internal monologue or silent soliloquy" and "talking to oneself in silence,"¹¹ and even, at one point, as "silent babblings."¹² He suggests that the basic genus is talking, the differentia being the amount of sound and conspicuous lip-movement produced:

The sealing of the lips is no part of the definition of thinking. A man may think aloud or half under his breath; he may think silently, yet with lip-movements conspicuous enough to be read by a lip-reader; or he may, as most of us have done since nursery-days, think in silence and with motionless lips. The differences are differences of social and personal convenience, of celerity and of facility.¹³

To be sure, Ryle differs from the Behaviorists in that he is not doing empirical psychology when he makes these claims. He would not be disconcerted in the least by the empirical discovery that thinking

can and does go on even when there is complete paralysis of all the organs and muscles involved in talking.¹⁴ Ryle's claim is merely a conceptual claim, namely that the *meanings* of expressions referring to thinking are to be explained in terms of the meanings of expressions referring to talking. To speak of "thinking" during complete paralysis would be to use that term in a marginal or extended way, Ryle would say.

The fact of the matter is that talking and thinking are quite different concepts. Talking necessitates public and overt behavior; thinking does not. Even people who "talk to themselves" make noise; Ryle's "talking to oneself in silence" is self-contradictory since "talking," taken literally, involves using the spoken word. What are "silent babblings"? Could a brook babble *silently*? One can make some sense of these as metaphors because there are similarities, to be sure, between talking and thinking. It is these similarities which allow such expressions as "think out loud" or "speak one's mind." But if I had to choose (I don't), I would rather say that talking is to be analyzed as a kind of thinking (out loud) or a giving expression to thoughts than put it the other way round, as Ryle would have it. In point of fact both assimilations are incorrect. A more desperate move might be to take thoughts to be *dispositions* to talk. But, as we have seen, thoughts are happenings, whereas dispositions are not. So this will not do either. At the very least, a thought must be taken to be the *acquisition* of a disposition to speak. But that is conceptually far removed from speaking, and allows for the possibility that thoughts are mental, in a sense which will, in the next section, be explained.

The "Asymmetry" of Mental Reports

Feelings and thoughts are *mental events* (although they are not the only mental events—e.g., images and dreams would also qualify). That they are events is clear from the fact that they occur at some datable time. But what is the force of the qualifying "mental"? Although we cannot pursue this matter very far, it does seem to be generally agreed by the bulk of philosophical thought on this matter that

¹⁰ Descartes, *Meditations* (II): "What is a thing which thinks? It is a thing which doubts, understands, conceives, affirms, denies, wills, refuses, which also imagines and feels."

¹¹ Ryle, *op. cit.*, p. 27.

¹² *Ibid.*, p. 58.

¹³ *Ibid.*, p. 34.

¹⁴ Patients with enough curare to effect complete muscular paralysis report afterwards that there was no loss of consciousness, sensations, ability to think or remember, image, etc. See S. M. Smith, H. O. Brown, J. E. P. Toman, and L. S. Goodman, "The Lack of Cerebral Effects of d-Tubocurarine," *Anesthesiology*, vol. 8 (1947), pp. 1-14.

feelings and thoughts as such are not overt, public, observable-in-principle happenings in the way that bodily happenings are overt, public, observable-in-principle. A crucial indication of this difference is found in the frequently noted "asymmetry" of first and third person reports. If John reports that just now it suddenly occurred to Charles that Hitler might still be alive, John makes this report on the basis of sense observations or other evidence and he can be mistaken because of mistaken observation or inference, whereas if John reports of himself that just now it suddenly occurred to him that Hitler might still be alive, John does not make this report on the basis of sense observation or evidence and so he cannot be mistaken because of mistaken sense observation or inference.

The question of "asymmetry" is still a matter of much controversy, especially because it is very easy to express it badly. For example, if I say that *the way* in which you know I am in pain is different from *the way* in which I know I am in pain (as John Wisdom sometimes puts it in his series, "Other Minds"), then it makes it look as if I have some *method* for coming to know I am in pain, which is absurd. Similar objections can be made to formulations using "my evidence for believing I am in pain," or "my grounds for believing I am in pain." Thus Ryle attacks a straw man when he attacks "Privileged Access" by pointing out:

The sorts of things that I can find out about myself are the same as the sorts of things that I can find out about other people, and the methods of finding them out are much the same. . . . In principle, as distinct from practice, John Doe's ways of finding out about John Doe are the same as John Doe's ways of finding out about Richard Roe.¹⁵

This misses the mark because the asymmetry is not one of different *methods* or *ways of finding out*. I do not have a way of finding out I am in pain, so *a fortiori* I cannot have a different way from yours. Better to put the asymmetry by saying that while you must have some way of finding out I am in pain (observing me, say), I just know when I am in pain. Even this last remark needs qualification, since I may be so engrossed in something that for a while I may forget my discomfort.¹⁶ This is something like the Lieutenant's forgetting he's an

officer during the baseball game, although if specifically asked he will of course know the answer. Thus I know when I am in pain in the sense that *if asked* I could answer correctly, although I may not be aware of the pain until I am asked. But I know without having to use my senses or appeal to evidence. This is one way to put this asymmetry which characterizes the mental.

Let me stress that this discussion of mental events is concerned with meanings of expressions and not with empirical facts. I have not shown that there actually are any mental events (although we all know that sensations, feelings, and thoughts do occur), nor what their relation, if any, to the body might be. At best I have held that expressions referring to feelings and thoughts are not *synonymous* with expressions referring to behavior or dispositions to behave. This is a conceptual or linguistic point, not a psychological one.

If, then, we look on the general goal of Ryle's analysis as that of dissolving the traditional mind-body problem, we must say that he has failed. For he himself ends up giving a special place to certain sorts of *mental episodes*. While it may be the case that the bulk of our expressions purporting to refer to mental episodes can be analyzed as referring to dispositions toward or exercises of overt behavior, still there do remain the important classes, thoughts and feelings, at the very least, which resist this analysis. And that represents the defeat of Ryle's general program. As Peter Geach says about this failure, "A logical principle allows of no exceptions—not even if the exceptions are events that only James Joyce would put into a novel."¹⁷

It may be worth noting one more recent attempt to give a physicalistic analysis of mental events. Stuart Hampshire examines feelings such as anger.¹⁸ He proposes that they be treated as *inclinations to act* in certain standard ways in certain standard situations. The weasel-word here is "inclinations." Hampshire realizes this, but makes only the negative point that he does *not* mean it in the merely hypothetical sense that certain behavior would occur if certain conditions were satisfied but in a categorical sense as "something that may occur at a certain moment, and may then immediately disappear, although it may also continue over a certain period of time."¹⁹ In this way

¹⁵ Ryle, *op. cit.*, pp. 155–156.

¹⁶ Gilbert Ryle, "Feelings," *The Philosophical Quarterly*, vol. 1 (1951), p. 200.

¹⁷ Peter Geach, *Mental Acts* (London, 1960), p. 5.

¹⁸ Stuart Hampshire, *Feeling and Expression*, Inaugural Lecture, University College (London, 1961), 22 pp.

¹⁹ *Ibid.*, p. 20.

Hampshire feels he has found a "middle way, which is neither a Cartesian dualism on the one hand, nor on the other hand a reduction of that which is distinctively mental to its overt behavioral expression."²⁰

Hampshire's "middle way" will not bear the weight of scrutiny. If his "inclinations" are things "that may occur at a certain moment," then they are not merely dispositions but actual events. Such "inclinations" to act in certain ways would go in the same class with occurrences such as hankerings, longings, yearnings, cravings, and itchings to act in certain ways. Since events such as these admit of the asymmetry of the first and third person, they meet the conditions for what I have called *mental events*. So we are still left with expressions which are not definable in purely physical terms.

It is possible that someone will come up with an analysis which makes it clear that mentalistic expressions are indeed synonymous with some sort of purely physicalistic expression. Although many have tried, no one has as yet succeeded in doing so. I take this to indicate that there does exist a real difference in meaning here. This is a quite modest claim. Mental events and physical events may still, in some sense, be "one," as we shall see when we come to the Identity Theory, in Section III. But they will not be one in the sense in which bachelors and unmarried men are one, viz., in the sense that the expressions, "bachelor" and "unmarried man" are synonymous.

Intentionality and the Mental

So far in this section I have been concerned with the mental side of the mind-body problem and have taken as the crux of this, mental *events*, in particular, feelings and thoughts. These will remain the paradigm cases throughout this paper. However, mention should be made of an alternative proposal for a paradigm of the mental, that of "intentionality."²¹ Intentionality characterizes those states such as wanting, hoping, wishing, seeking, believing, and assuming, where something may be the object of that state whether that something actually exists or not. For example, a person may want a dog that can talk. This talking dog

may not exist yet it is a fact that a person may want it, seek it, and believe in it. It is, though it may not be that merely, a "product of his mind." The ability to create such products is held by some to be sufficient if not necessary for something's being mental. It is the thesis of intentionality that intentional states are not analyzable in physical terms,²² and hence that they are essentially mental phenomena.

For our purposes we do not need to decide whether to accept or reject the thesis of intentionality. We need only note that it is not pertinent to the mind-body problem here discussed, namely the problem of the relation between mental *events* and bodily events. Even if intentional phenomena do represent a kind of irreducibly mental phenomena, they are neither necessary nor sufficient for the occurrence of mental *events*. A person may believe that there are tigers in Asia without ever explicitly thinking this belief on some actual occasion; for example, he may believe that there are tigers in India and that India is in Asia, without ever having thought explicitly that there are tigers in Asia. Hence, intentionality is not a sufficient condition for a mental event. And since there is nothing intentional about a sudden bout of nausea or "free floating" anxiety, intentionality is not a necessary condition of a mental event. There are, of course, important connections between intentional phenomena and mental events. For one thing, one important sort of mental event, what I have called thoughts, are basically intentional. For another, whenever intentional phenomena take the form of events they will be *mental events*, as for example in the occurrence reported by "Suddenly for a moment he really believed he would be well again." Finally the two often go together; when we want something or believe something we may have certain sensations, feelings or thoughts which are symptoms or concomitants of the want or belief. So intentional phenomena and mental events are closely related. Nevertheless for our purposes we shall not be concerned with the intentional as such. However, when we come to discuss machines, we shall consider briefly the question whether intentional phenomena could be attributed to machines (p. 90).

²⁰ *Ibid.*, p. 21.

²¹ A leading contemporary supporter of this view is Roderick M. Chisholm. See "Intentional Inexistence," in *Perceiving* (New York, 1957) Chap. 11.

²² This thesis is attacked by Wilfrid Sellars, "Physical Realism," *Philosophy and Phenomenological Research*, vol. 15 (1954), pp. 13-32. For a (somewhat inconclusive) discussion between Sellars and Chisholm, see "Intentionality and the Mental," *Minnesota Studies in the Philosophy of Science*, II ed. Herbert Feigl *et al.*, pp. 507-539.

Summary

We can summarize this section by saying that, yes, Virginia, there is a mind-body problem. That is, there are genuine events which do occur at datable times and which are not describable merely in physical terms. It is Ryle's achievement to have shown that these events are not as widespread as some have thought, and specifically to have shown that much of our intelligent and purposeful activities can go on without them. Nevertheless these events do occur. Ryle fails in his attempt to deny them a special status, which they deserve, because of their feature of "privileged access," the feature by virtue of which the person to whom the events occur knows they occur without having to make the observations or inferences others must make to know of their occurrence. Now given these events with their special status the question then arises how these events are related, if at all, to events which are describable in physical terms. This is the mind-body problem.

II. MACHINES AND THE MIND-BODY PROBLEM

The Claim of Cybernetics

In recent times much attention has been devoted to the very close analogy between the workings of the mind and the workings of certain sorts of mechanical devices. Two sorts of mechanical devices have been especially important: servomechanisms and computers. Servomechanisms can achieve and maintain some predetermined state by use of feedback; here we have something very similar to purposeful human behavior. Computers can start with a set of characters and finish with a new set of characters by performing sequences of operations; here we have something very similar to human reasoning. And in purposeful behavior and reasoning we have features which have been traditionally taken to be at the very heart of the mental. So it has been argued by many that the solution to the mind-body problem is to be found in the consideration of such machines. The neurologist, W. S. McCulloch, has proposed that we "conceive of the knower as a computing machine"²³ and Hilary Putnam concludes that if we do so we shall see that "it is no longer possible to believe that the mind-body problem is a genuine theoretic-

cal problem."²⁴ This is so, he argues, because "every philosophical argument that has ever been employed in connection with the mind-body problem . . . has its exact counterpart in the case of" computing machines, which shows, Putnam believes, that the problem is "purely verbal" and therefore "obviously of no importance."²⁵ In this section I propose to examine this claim.

A number of questions must be distinguished here. First, there is the question how close is the analogy. That there is only an *analogy* here must be admitted from the start. Even if we could take an enormous number of living cells (or even sub-cellular components) and combine them so as to get an exact replica of a human, there would be at least the difference of origin. But a more important difference is that of composition; the basic stuff of which humans are made, living organic material, is different from the stuff of which any machine, built or contemplated, is made. So at best there can be only a similarity. Furthermore, so far as design is concerned, while we can say very little about the design of the brain, it still seems reasonable to say that the design of the brain is different from that of any machine, built or contemplated. So far as functioning is concerned, it is quite likely that there will always be differences here, such as speed in handling certain sorts of problems, if only as a consequence of the differences in material or design. How are we to decide how close the analogy is?

To begin with, what sort of machine do we mean? First, we are thinking of something *artificial*, something deliberately designed and made by someone. Second, to eliminate possible confusions, we are thinking of something composed largely of inorganic material, certainly with no living organic components. Third, we are concerned with what present theory indicates is *possible* in the way of machines, not what is practicable or feasible. It does not matter for our purposes that a machine analogous to a human in some respects would never be built because it would cost too much, take too long to build, would be too large, break down too often, etc. Such problems are merely technical in nature. Consider the problem of memory storage, for example. The early electronic machines could store only some couple of hundred items, whereas present machines are able to store billions of items. This may still be well below the memory capacity of the human brain

²³ "Through the Den of the Metaphysician," *British Journal for the Philosophy of Science*, vol. 5 (1954); see p. 19.

²⁴ "Minds and Machines," *Dimensions of Mind*, ed. Sidney Hook (New York, 1961), pp. 160-161.

²⁵ *Ibid.*, pp. 160-161.

but only technological ingenuity is needed to increase machine memory storage. We do not have here a problem of principle; in principle there are no limits to the size of a computer memory.

In our comparison, then, of machines and human beings, and, in particular, human brains, we start with a difference in origins and composition. How about design? How similar are they so far as structure and internal functionings are concerned? The fact of the matter is that we do not know. We are still very ignorant about the actual structure and functioning of the human nervous system, especially the central nervous system. So there is very little knowledge at present concerning the degree to which the design of a computer is analogous to that of the brain. Computer theory may be a source of speculations concerning brain mechanisms, just as information about the brain may provide suggestions for innovations in the design of computers. But so far as the actual similarity of design is concerned, very little is known at present. The hunch of some neurophysiologists is that they are remarkably similar in many respects. But this is just a hunch.

Turing's Question

In trying to give some precision to the question of the extent of the analogy, A. M. Turing raised the following question: Could there be, in theory, a computer which would so answer questions put to it that the answers would be indistinguishable from those a man would give?²⁶ Here the criterion concerns not the origin, composition, or design of the machine but its capacities, or, if you like, its behavior, or, if you wish to be very strict, its "output."

Much of the discussion in recent times has concerned the question whether there are any discernible differences in the output of humans and possible machines. One stumbling block has been that many descriptions of human activities make no sense at all when applied to nonliving systems, e.g., giggling. Probably most of the expressions currently used to describe the output of machines, "computing," "receiving information," "remem-

bering," "adding and subtracting," "making deductions," "playing chess," "writing checks," etc., involve a shift of meaning when applied to machines. But it is a perfectly natural shift in these cases (whereas I find it difficult to imagine a computer contemplated at present which might be intelligibly said to be able to giggle). As Turing pointed out with reference to the applicability of the word, "think," to computers:

The original question "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.²⁷

So far as machine performance is concerned, there have been extraordinary achievements in the last decade. Engineers report that "at present, we have or are currently developing, machines that prove theorems, play games with sufficient skill to beat their inventors, recognize spoken words, translate text from one language to another, speak, read, write music, and learn to improve their own performance when given training."²⁸

Have we reached the point yet where the answers a machine would give to questions put to it would be indistinguishable from human answers? Turing's question has already prompted some empirical testing. For example, a machine has been programmed which will conduct an intelligent conversation about the weather.²⁹ The machine program is comparatively small, containing about 800 instructions, yet the machine replies to most remarks or questions about the weather appropriately and intelligently. Occasionally, but surprisingly infrequently, the machine generates replies that do not make sense. An experiment was performed to see how close the computer came to simulating ordinary conversation.³⁰ Twenty-five comments and questions taken more or less at random from a large number of actual conversations about the weather were fed into the machine. The reply of the machine to each comment was mixed with the responses of nine ordinary people to the comment and a different group of people

²⁶ A. M. Turing, "Computing Machinery and Intelligence," *Mind*, vol. 59 (1950), pp. 433-460.

²⁷ *Ibid.*, p. 442.

²⁸ T. Marill, "Human Factors in Electronics," *Transactions of the Institute of Radio Engineers*, vol. 2 (1961); as quoted in Ulric Neisser, "The Imitation of Man by Machine," *Science*, vol. 139 (1963), p. 193.

²⁹ This is reported at length in Edmund C. Berkeley's *The Computer Revolution* (New York, 1962); see pp. 87-110.

³⁰ The experiment was carried out by Patrick J. McGovern and reported in *Computers and Automation*, Sept. 1960. The gist of this paper is in Edmund C. Berkeley, *ibid.*, pp. 102-110 and 203-210.

was asked to judge which of the ten replies to each comment was produced by the computer. If they had been just guessing at random they would have gotten 10 per cent right. Actually they got 16.8 per cent right. This indicates that most of the responses of the machine were indistinguishable from those generated by intelligent humans. And it would not have been difficult to add to the program so as to improve the performance of the machine.

Conversations about the weather do not require *very much* intelligence (hence their frequency), but we do see here the possibility of constructing programs for conducting more profound conversations. It is such achievements as this which lead many experts today to think that in a few years it will be possible to construct a machine such that "it will be impossible for a human being in another room to tell whether he is conversing with a computer or with a human being."³¹ This was what Turing was inclined to say,³² and many contemporary philosophers would agree with this.³³

How about emotions and motives? Many have argued that it is in this area that the crucial differences are to be found:

People get bored; they drop one task and pick up another, or they may just quit work for a while. Not so the computer program; it continues indomitably.³⁴

But if we remember that at this point in our discussion we are concerned only with machine *output*, with answers to questions that are put to the machine, then it is reasonable to think that machines can be programmed to exhibit boredom, restlessness, shifts in attention, sulkiness, rage, etc.³⁵ Could a machine make jokes? Here one seems to come to something so characteristically human that it is hard to imagine a machine able to do it. But this arises, I think, from the fact that we do not

understand very well what humor is. If machines can write tolerable poetry and compose tolerable music, why should they not be able to make jokes also? But could a machine *enjoy* poetry or jokes? Could it *feel* boredom or rage? With these questions we go beyond the Turing formulation in terms of *output* to questions of internal states. We shall return to these questions shortly.

The Objection from Gödel's Proof

The most important source of objections to Turing's thesis is the fact that computers must be *programmed*, i.e., must be told what to do. This means that they can do only what can be programmed and what they individually are programmed to do. And this means that for any machine there will of necessity be things outside the machine's program which it cannot do and, perhaps, if anything is unprogrammable, some things which no machine could ever do. We might call this the "Argument From Design."

A particular application of this argument comes from considerations of Gödel's theorem. Turing was one of the first to make this application.³⁶ The matter is put thus by J. R. Lucas:

Gödel's theorem states that in any consistent system which is strong enough to produce simple arithmetic, there are formulae which cannot be proved-in-the-system, but which we can see to be true. . . .

Gödel's theorem must apply to cybernetical machines, because it is of the essence of being a machine, that it should be a concrete instantiation of a formal system. It follows that given any machine which is consistent and capable of doing simple arithmetic, there is a formula which it is incapable of producing as true—i.e., the formula is unprovable-in-the-system—but which we can see to be true. It follows that no machine can be a complete or adequate model of the mind, that minds are essentially different from machines.³⁷

³¹ Edmund C. Berkeley, *ibid.*, p. 110.

³² A. C. Turing, *op. cit.*, p. 442.

³³ The following articles support this contention: D. M. MacKay, "Mindlike Behavior in Artifacts," *British Journal for the Philosophy of Science*, vol. 2 (1951), pp. 105-121; W. Ross Ashby, "Can a Mechanical Chess-Player Outplay Its Designer?" *ibid.*, vol. 3 (1952), pp. 44-57; D. M. MacKay, "Mentality in Machines, III," *Proceedings of the Aristotelian Society, Supplementary Volume XXVI* (1952), pp. 61-86; Michael Scriven, "The Compleat Robot," *Dimensions of Mind (op. cit.)*, pp. 113-133; Hilary Putnam, "Minds and Machines," *ibid.*, pp. 138-164.

³⁴ Ulric Neisser, "The Imitation of Man by Machine," *Science*, vol. 139 (1963), p. 194.

³⁵ See Silvan S. Tomkins and Samuel Messick, *Computer Simulation of Personality* (New York, 1963).

³⁶ A. M. Turing, "On Computable Numbers, with an Application to the Entscheidungs-problem," *Proceedings of the London Mathematics Society*, vol. 42 (1937), pp. 230-265. See Turing's later article, "Computing Machinery and Intelligence," *op. cit.*, pp. 444-445, for his own reply.

³⁷ J. R. Lucas, "Minds, Machines, and Gödel," *Philosophy*, vol. 36 (1961), pp. 112-113. A more cautious position is maintained by Ernest Nagel and James R. Newman in their *Gödel's Proof* (New York, 1958) that Gödel's results show the superiority of the human brain over "currently conceived artificial machines" (see pp. 100-101).

Hilary Putnam argues that such a line of reasoning:

is a misapplication of Gödel's theorem, pure and simple. Given an arbitrary machine, T , all I can do is find a proposition U such that I can prove

(3) if T is consistent, U is true

where U is undecidable by T if T is in fact consistent. However, T can perfectly well "prove" (write down a proof of) (3) too! And the statement U , which T can't "prove" (assuming consistency), I can't prove either! (Unless I can prove T is consistent, which is unlikely if T is quite complicated.)³⁸

So, Gödel's theorem does not show the essential superiority of man over machine. One might think that Putnam's last qualification weakens his case, for there may be a plausible formal proof for the consistency of T in a richer system. But so long as we speak of *formal* proofs, it is within the competency of a suitably constructed machine to provide such proofs.

This, however, does not deal directly with Lucas' point that U is "unprovable-in-the-system—but which we can see to be true." After all, strictly speaking, what can be proved, by us *and* machine, is not Putnam's proposition (3) but

(3') if T is consistent, U is *undecidable*

Now because of the peculiar way in which Gödel constructs U , we can see that it says of itself that it is undecidable, so if U is undecidable, we can see that it is true, but Lucas might argue that there is no reason to think that the machine could see this. However, such an argument fails. Because of the way U is constructed, the proposition that U is undecidable and the proposition that U is true are one and the same proposition. There is no difference between seeing that U is unprovable if the system is consistent and seeing that U is true if the system is consistent. So no difference has so far been established between what a machine can do and what we can do.

To return to the general Argument From Design of which the Gödel theorem is a special case, it is

argued that the machine's program establishes not only what it can do but also what it *cannot* do; therefore every machine is forever limited in what it can do, whereas the brain is not so limited and therefore essentially different from a machine. I do not find this argument convincing. First, it is reasonable to think that the brain, too, has its limitations; so the difference is likely to be only one of degree, not of principle. Second, there have already been enormous gains in machine flexibility, particularly in terms of machines with built-in randomizers and machines which modify their own programs through experience. Especially promising is the development of machines with "heuristic" devices, at present, typically, rules of thumb for taking shortcuts.³⁹ These rules of thumb not only save the computer time and enable it to come up with more elegant solutions, but also allow it to deal with problems whose solution cannot be reached in a finite number of steps. At present such heuristic devices are built into the machine, but in principle a machine could be designed which would discover new ones for itself through experience. So we see how a machine might come to transcend some of its original limitations. Thus the Argument From Design only works for comparatively primitively designed machines. As F. H. George points out:

In cybernetics we are not dealing with machines that are wholly specified in advance. They are self-programming or self-organizing.⁴⁰

Such machines can learn from past experience, extrapolate into the future, come up with novel configurations (for example, via the randomizer), discover new proofs, and make evaluative judgments (for example, via the heuristic devices). These machines show the inappropriateness of the standard charge that machines are incapable of genuine creativity. There are already chess-playing machines which come up with novel strategies on the basis of games already played and theorem-proving machines which come up with better proofs than Euclid or Russell-Whitehead

³⁸ Hilary Putnam, Review of *Gödel's Proof* by Nagel and Newman, *Philosophy of Science*, vol. 27 (1960), p. 207. See also Putnam's "Minds and Machines," *op. cit.*, p. 142. For a reply by Nagel and Newman, see *Philosophy of Science*, vol. 28 (1961), pp. 210–211.

³⁹ A. Newell, J. C. Shaw, H. A. Simon, "Report on a General Problem-Solving Program," *Information Processing*, UNESCO, 1960, pp. 256–269; H. Gelernter, "Realization of a Geometry Theorem Proving Machine," *ibid.*, pp. 273–282. In the latter the machine uses as a heuristic a *diagram* which enables it to test various purported solutions and reject those which are unlikely to work.

⁴⁰ F. H. George, "Minds, Machines, and Gödel: Another Reply to Mr. Lucas," *Philosophy*, vol. 37 (1962), p. 62. See also J. J. C. Smart, *Philosophy and Scientific Realism* (London, 1963), pp. 116–120.

ever provided. In the face of such advances, how can we say with any confidence that such-and-such tasks are forever beyond the machine? In point of fact we cannot.⁴¹

Let us consider briefly the question of "intentionality" in machines (see p. 85). So far as output is concerned, intentionality may be ascribed to the machine. For example, a machine may look (in its memory banks) for tigers in India and, on the basis of its information, come to believe that there are or are not such things. If ascribing intentionality to machines involves an extension of meaning, this is because at present our linguistic rules allow us only to make such ascriptions to things which have mental events. Whether machines could have mental events is a question we shall come to shortly.

With respect to *output*, then, I am inclined to believe that the similarity between man and machine is very close, at least in principle and in relevant respects and, specifically, that there is no human output which is not duplicatable by a machine. This is, of course, merely an *empirical* hypothesis. There is a constant temptation to argue that since the brain can do something and the brain must have a design, follow rules, or have a "program," it follows that a machine could do it too. But this is to beg an empirical question. It is at least logically possible that the brain does not function in any rule-like way. So the fact that a brain could do it does not show *a priori* that a machine could also. What does make it likely that a machine could also do it is the recent advances in machine theory and technology.

One must not exaggerate the consequences of this output similarity. For example, what does it tell us about the *internal* similarities of brain and machine? Absolutely nothing. We know for a fact that similarity of output does not show similarity of origin or of composition. So far as the design of the brain is concerned, it might well be (probably is) entirely different from that of any known computer. There is no reason to think that how a computer goes about adding or playing chess is anything like how the brain operates in these matters. All we know is that so far as the *output* is concerned, that of an adding machine or a chess-playing machine is, in principle, probably indistinguishable from that of a human. Since we know so little about the internal processes of the brain,

computer theory may give us a fruitful source of hypotheses, and these hypotheses may become well confirmed. But as yet very little is established in this area.

The Machine and Mental Events

If performance were the logically decisive factor in determining whether something did or did not have a mind or consciousness, then there would be little question that, in theory at least, we could build machines with consciousness. There would still be important differences, to be sure, certainly in origin and composition and probably in design. But such differences are not relevant to whether something was an intelligent or purposive performance. When we say that something was done intelligently or stupidly, on purpose or by accident, we do not commit ourselves to what the immediate origin of the thing was or what it was composed of or what its internal structure was. So if output were all that mattered, there would be no difference in principle between man and machine.

However, there is a further item of crucial importance, *mental events*. Could machines have sensations, feelings, or thoughts? Could a machine feel an ache or a smart; could it feel remorseful, proud, or angry; could it suddenly dawn on a machine that . . . or could a machine have the experience of realizing that . . . ?

Of course a machine could be programmed to *simulate* the having of mental events. It could print out reports on its inner condition ("Tube 37541 just blew out") or its progress in solving a problem ("Just thought up a well-formed formula which is neither an axiom nor an already proven theorem."). Or it might be programmed to print out such reports only on demand; otherwise it would keep its internal state to itself. Still, this is only simulation, only similarity of *output*. The question would remain whether the machine actually had feelings and thoughts in the sense in which we do.

It is tempting to look for some well-known feature of the machine and *identify* that with mental events, so that it becomes obvious that machines have mental events. Hilary Putnam argues that mental states in the machine are identical with what he calls the "logical states" of the machine.⁴² Any machine is composed of various pieces of

⁴¹ For a recent discussion and rebuttal of various alleged disabilities of machines, e.g., that they cannot reproduce, do the unpredictable, create, learn, understand, abstract, perceive, or feel, etc., see Michael Scriven, *op. cit.*

⁴² Hilary Putnam, "Minds and Machines," *op. cit.*

hardware, and as it operates the hardware goes through a succession of states; Putnam calls them "structural states." But the machine is also going through a succession of states in another sense. Its operations can be described in an abstract, logical way; the machine table or program gives such a description. Symbols in various locations may be replaced by symbols from other locations and in this way a sum may be worked out. Physically any number of structural states may serve to realize this logical operation. Putnam argues that the mind-body problem in the machine is simply the problem of the relation of the logical states to the structural states of that machine.

Illuminating as this is, it does not remove our further question whether machines have mental events. We can see this plainly if we apply Putnam's distinction to a human being, solving an arithmetic problem at the blackboard, for example. The structural states will be physical states of his body (and, in this case, the resulting changes on the blackboard, etc.). The logical states will be the adding, subtracting, etc., and the sub-operations which they require, carrying certain numbers from one column to another, moving decimal points, etc. All of these logical states could be specified in such a way that a machine or person could work out the problem by "mechanically" following the steps. However, there would still be the question: What *mental events* occurred? Did the person doing the sum have any thoughts while doing it? Perhaps, if very practiced, he had no thoughts at all as he worked out the problem. Perhaps he was thinking what a boring problem it was. Perhaps he was thinking the rules ("carry the three," "drop the decimal point," etc.) as he did the operation. So in addition to Putnam's "structural" and "logical" operations there could also occur mental operations, in the sense in which I have been talking of mental events. Our question, so far as machines are concerned, is precisely whether there might not be these additional mental events.

Some philosophers have tried to dismiss such questions as meaningless, but it seems to me that they are quite meaningful. If you are inclined to think that it is a contradiction to speak of a *machine* having feelings or thoughts, then we can reformulate the question to ask whether there might be things just like people in all respects except that they were made in the electronics laboratory. Imagine computers getting better and better at

more and more, as theory and technology improve, until (in principle at least) they behave just like people. Might they not have mental events then too? Could not an extremely complex organization of transistors, tapes, magnetic cores, photoelectric cells, etc., turn out to have sensations, feelings, and thoughts just like ours? Is this not a real possibility? Some are inclined to rule this out:

A machine made out of vacuum tubes, diodes, and transistors cannot be expected to have consciousness. I do not here offer a proof for this statement, except that it is obvious according to well-disciplined common sense.⁴³

But I know of no argument to show this.

The crucial question is how we could tell. Here we have the Other Minds problem with a vengeance. So far as behavior is concerned, our contemplated machine's output matches that of a human. However, its constituents and design are utterly different from that of a human. Does it have mental events or not? If it does, we now have the analog of the mind-body problem, the mind-machine problem: How are the mental events related to the machine? Are they mere concomitants, or mere by-products, or able to produce changes in the machine? We have exactly the same possibilities as in the traditional mind-body problem, and the same considerations are relevant in both cases. For example, to anticipate our discussion of Interactionism in Section IV, suppose it turns out that our machine occasionally comes up with solutions to problems and we cannot explain in terms of machine mechanics how it came up with these solutions, but we can explain if we hypothesize causally efficacious nonphysical events. Could it not happen that in the case of certain sorts of problems the machine idles for a while and then suddenly starts printing out a solution? We might say there was a mechanical explanation which we had not yet found or we might say that there was a randomness in the machine which resulted in a lucky outcome. But might we not also speculate that the machine had had insightful thoughts which produced the solution? Might there not be circumstances such that this speculation would offer the best explanation?

Since there is no difference in principle between the mind-body problem and the mind-machine problem, we shall postpone discussion of these issues in the next two sections. The relevant question here is still how we could tell whether machines have

⁴³ Satoshi Watanabe, "Comments on Key Issues," *Dimensions of Mind* (op. cit.), p. 136.

mental events at all. I shall discuss only one recent attempt to answer this, the ingenious suggestion of Michael Scriven that we so design our computer so that it cannot tell lies and then simply *ask* it if it has mental events. The problem is how to put our question to the machine. This is done by feeding into the machine both our ordinary language of mental predication and also behaviorist language so that it can distinguish as well as we can between a person's being in pain as opposed to merely behaving as if he were. Then:

having equipped it with all the performatory abilities of humans, fed into its banks the complete works of great poets, novelists, philosophers, and psychologists, we now *ask* it whether it has feelings. And it tells us the truth since it can do no other.⁴⁴

The trouble with this is that it begs the question by assuming that the machine will understand the question in the sense we intend. Might it not take the distinction between having a mental event and merely acting as if one did as the distinction between two sorts of behavior, the former more consistent and extensive and the latter less consistent or extensive? That would not be *our* concept. Our concept involves more than a certain sort of behavior; to understand the more involved requires having had the experience itself, or an analogous one at least. So whether the machine understands our question in our sense or not itself depends upon the prior question whether it has feelings itself. If it has not had feelings then it does not even understand what *we* mean when we ask it if it has feelings. The real question is whether the machine could understand our question. Scriven's reply to objections of this sort is that "we have every good reason for thinking that it does understand, as we have for thinking this of other *people*."⁴⁵ But this is not true since machines would have the further differences from people of origin, composition, and, perhaps, design. And these differences make the Other Minds inference much tougher in the case of machines.

One possible way to find out, although almost too farfetched to consider, would be to replace various parts of one's own brain with computer parts and see what happens. One might find oneself displaying pain behavior without feeling pain. But even here, the outcome might be inconclusive because the fact of only *partial* replacement might make all the difference. At present we do not know

enough about the brain even to imagine plausible experiments.

The upshot of this is that, if you were able to make a machine as complex as a human, then: (1) that machine *might* have the sorts of occurrences I have called mental events, and (2) if it did have mental events we might never be able to say with confidence whether it did or not. Now (1) leads us to the mind-machine problem, i.e., how we are to conceive the relation between machine and its mental events. And (2) is the problem of other minds for machines. How ironical it is that we should be led into agreeing with Putnam's principle that "every philosophical argument that has ever been employed in connection with the mind-body problem . . . has its exact counterpart in the case of" computing machines. He concluded from this that the mind-body problem was purely verbal whereas we have concluded from the same principle that the mind-machine problem is of genuine factual significance.

Summary

So far as output is concerned there is no reason to believe that humans are capable of doing things which machines could not be capable of doing. If we are Behaviorists, then we shall hold that machines are just as intelligent, purposive, and conscious as humans. If we believe, as I do, that there still remains the non-behavioral question whether machines could have mental events, then we must admit that here is a genuine empirical issue which may never be settled. At any rate it is very difficult to imagine what kind of experiment or observations might cast any light on this question. This does not make it any less significant a question.

III. THE IDENTITY THEORY

The Theory

In Section I, it was maintained that mental-event terms are not definable in physical terms. What follows from this indefinability? In the past, many have assumed that it was sufficient to prove the need for some sort of *dualistic* theory of mind and body—either a dualism of substances or, at the very minimum, a dualism of events. For, it was argued, if the terms were different in meaning, then

⁴⁴ Michael Scriven, *op. cit.*, p. 132.

⁴⁵ *Ibid.*, p. 133. Author's italics.

mental events were one thing and physical events and dispositions quite another. A typical line of argumentation is the following, by J. B. Pratt:

To say that consciousness is a form of matter or of motion is to use words without meaning. The identification of consciousness and motion indeed can never be refuted; but only because he who does not see the absurdity of such a statement can never be made to see anything. . . . If he cannot see that, though consciousness and motion may be *related* as intimately as you please, we *mean* different things by the two words, that though consciousness may be *caused* by motion, it is not itself what we mean by motion any more than it is green cheese—if he cannot see this there is no arguing with him.⁴⁶

In pointing out the linguistic point concerning difference of meaning, Pratt thinks he has refuted any sort of monistic theory, and so he then goes on to discuss various dualistic theories such as Epiphenomenalism, Interactionism, and Parallelism. But there is an error here. The author has overlooked a variety of monistic theory which has recently been given explicit formulation and discussion. It is called the Identity Theory.⁴⁷

Identity Theorists appeal to the standard logical distinction between *intension* (or meaning or significance or connotation) and *extension* (or reference or denotation). Agreeing that the intension or meaning of mental and physical terms are different, they hold the *empirical* thesis that there turns out to be identity of extension or referent, that is, that the actual events referred to by mental terms turn out to be the very same as those referred to by certain physical terms. To take an analogy, if we consider the expressions, "duly elected president of the United States" and "commander-in-chief of the United States armed forces," we will agree that their intensions or meanings are quite different but yet the actual person referred to by the one turns

out to be the very same person as that referred to by the other expression. Further examples of difference in meaning with identity of referent are "human being" and "featherless biped," "the Evening Star" and "the Morning Star," "water" and " H_2O ," and "lightning" and "electrical discharge (with further specifiable properties)." In all these cases, it turns out as a matter of empirical fact that the actual things referred to by the various pairs of expressions turn out to be one and the same. Mere semantic analysis of meanings would not tell you this; one must go out and look. With respect to the Identity Theory, it is maintained that expressions such as "pang of grief," "sudden thought that today is a holiday," and "feeling of pain in the left thumb" are indeed different in *meaning* from expressions referring to physical events but will (probably) turn out to be identical in reference with expressions referring to physical events. In short, it is held that mental events are physical in fact, although the words we use to refer to them differ in meaning.

The Identity Theory rests on an empirical hypothesis. It hypothesizes that for each particular mental event there is some particular physical event which always occurs and is such that whenever that physical event occurs then the mental event occurs. (The theory proposes to explain this by the further hypothesis that the physical event and the mental event turn out to be one and the same event.) So far as this empirical hypothesis is concerned, as yet very few details are known. The relevant physical events presumably occur in the nervous system, probably in the central nervous system (although peripheral feedback may be necessary). How much more localization there is remains uncertain. The Identity Theory would be empirically false if, to take an extreme, mental events continued to occur despite destruction of the entire brain matter (as is

⁴⁶ J. B. Pratt, *Matter and Spirit*, 1922; reprinted in Maurice Mandelbaum, Francis W. Gramlich, and Alan Ross Anderson, *Philosophic Problems* (New York, 1957), p. 266.

⁴⁷ For some recent statements and defenses of the Identity Theory, see: U. T. Place, "Is Consciousness a Brain Process?" *British Journal of Psychology*, vol. 47 (1956); Herbert Feigl, "The 'Mental' and the 'Physical'," in H. Feigl, G. Maxwell, and M. Scriven, eds., *Concepts, Theories, and the Mind-Body Problem* (Minneapolis, University of Minnesota Press, 1958); J. J. C. Smart, "Sensations and Brain Processes," *The Philosophical Review*, vol. 68 (1959); Hilary Putnam, "Minds and Machines," in Sidney Hook, ed., *Dimensions of Mind* (New York, 1960).

For some recent criticisms of the Identity Theory, see: George Pitcher, "Sensations and Brain Processes: A Reply to Professor Smart," *Australasian Journal of Philosophy*, vol. 38 (1960); J. T. Stevenson, "Sensations and Brain Processes: A Reply to J. J. C. Smart," *The Philosophical Review*, vol. 69, (1960); Kurt Baier, "Smart on Sensations," *Australasian Journal of Philosophy*, vol. 40 (1962); James W. Cornman, "The Identity of Mind and Body," *The Journal of Philosophy*, vol. 59 (1962). For replies by J. J. C. Smart to the above criticisms, see: "Sensations and Brain Processes: A Rejoinder to Dr. Pitcher and Mr. Joske," *Australasian Journal of Philosophy*, vol. 38 (1960); "Further Remarks on Sensations and Brain Processes," *The Philosophical Review*, vol. 70 (1961); and *idem* "Brain Processes and Incommensurability," *Australasian Journal of Philosophy*, vol. 40 (1962). For the present author's views on the Identity Theory, see: "Could Mental States Be Brain Processes?" *The Journal of Philosophy*, vol. 58 (1961); and "Mental Events and the Brain," *The Journal of Philosophy*, vol. 60 (1963).

claimed to occur by religious believers in disembodied life and the immortality of the soul). So the Identity Theory is, in principle at least, refutable, and is therefore a meaningful empirical theory.

There is a nonempirical side to the Identity Theory, however. For it adds to the empirical hypothesis that mental events have necessary and sufficient brain events the further proposal that those mental events be treated as *identical* with those brain events. Mental events are not conceived as some *extra* events causally produced by and fully dependent upon the brain events; they are to be conceived as one and the same as those brain events.

The Identity Theory has a 'deceptive air of simplicity about it. After all, we understand perfectly well some of the identities with which it is compared. When we say the Morning Star and Evening Star are identical, we mean that both names refer to one physical object at different times in different places. When we say the president and commander-in-chief of the armed forces are identical, we mean that both roles are assigned, by law, to the same person. What do we mean in the case of the Identity Theory? Do we mean (1) that certain *brain states are mental*, that they can be directly known by introspection, that someone without the slightest training in neurology can know without observation that certain incredibly complex events are going on in his infero-temporal cortex? Do we mean (2) that *mental events are physical*, that the thought that today is a holiday, for example, has a shape, a size, a charge, or a color, that it can be photographed, or perhaps smelled? Both of these interpretations seem most paradoxical.

An Idealistic Version of the Identity Theory

At least one Identity Theorist, Feigl, seems to suggest that the basic and underlying reality of these mental-physical unities is *mental*:

Speaking "ontologically" for the moment, the identity theory regards sentience (qualities experienced, and in human beings knowable by acquaintance) and other qualities (unexperienced and

knowable only by description) the basic reality. . . . It shares with certain forms of idealistic metaphysics, in a very limited and (I hope) purified way, a conception of reality.⁴⁸

Thus, although mental events are not the only reality, as Idealism would have it, they are the basic realities to which mental expressions refer, and neurophysiological expressions turn out to refer to these very same mental events. In Feigl's words, "according to the identity thesis the directly experienced qualia and configurations are the realities-in-themselves that are denoted by the neurophysiological descriptions."⁴⁹ Expressions which purport to refer to brain happenings turn out to refer, basically, to nonphysical, mental happenings!

This desperate attempt to avoid dualism is utterly implausible. Neurophysiological descriptions surely describe events, processes and states of the public, physical, spacially extended brain, and the criteria for applying such descriptions are public and physical. How could any conceivable discovery show, as Feigl would have it, that some of these descriptions really refer not to physical events primarily, but to "experienced qualia" or, as he also calls them, "raw feels."⁵⁰ After all, presumably something is going on neurologically when mental events occur and it is only reasonable to think that the neurological descriptions of those events refer to those events. Feigl would have it that when we think or feel, nothing of a neurological nature occurs at all, except in the trivial manner-of-speaking sense that we may use neurophysiological descriptive expressions, but these expressions refer not to neurophysiological happenings but to mental happenings. How would we know what neurophysiological descriptions to apply if the basic phenomena are mental and not physical at all?

In point of fact, Feigl presents us with a (thinly veiled) Interactionist Theory. He insists on "the efficacy of mental states, events, and processes in the behavior of human (and also some subhuman) organisms,"⁵¹ but on the official account of the Identity Theory, this efficacy arises solely from the (correlated) neurological events. Yet on the *sub*

⁴⁸ Herbert Feigl, "The 'Mental' and the 'Physical'," *Minnesota Studies in the Philosophy of Science*, vol. II, p. 474. Feigl goes on to say that he agrees with materialism in its rejection of irreducibly teleological phenomena. This is in the area of "intentional" phenomena, which we have mentioned but do not concern ourselves with in this paper. For a later, briefer, but essentially similar treatment by Feigl, see his "Mind-Body, Not a Pseudoproblem," in *Dimensions of Mind* (*op. cit.*), chap. 2.

⁴⁹ Feigl, *op. cit.*, p. 457.

⁵⁰ *Ibid.*, pp. 385, 416, 475, etc.

⁵¹ *Ibid.*, p. 388. Author's italics. See also p. 475.

rosa Idealism, only the mental event remains to be causally efficacious.

Materialist Versions

Whereas Feigl's Identity Theory dissolves, via an Idealism, into a conventional Interactionist view, other Identity Theories move in the opposite direction toward conventional Materialism. For them the underlying reality is physical throughout; the basic referent of both neurological and mental expressions is the brain. Mentalistic expressions form what is merely a different language for talking about physical events. The problem now becomes that of explaining the relation between the two languages.

It is generally agreed that mentalistic and physicalistic languages are not alternative languages for talking about the same things in the way that, say, French and English are. For in the case of French and English, suitably chosen pairs of expressions *mean the same*, are synonymous. But the Identity Theory does not hold, and indeed would be foolish to hold, that certain mental and physical expressions are synonymous. So this analogy is out.

B. A. Farrell suggests that mentalistic language involves an irrational, unempirical theory carried over from a primitive and superstitious past. He claims that notions like that of "mental event," as used in this paper, "can be shown to resemble an occult notion like 'witchcraft' in a primitive community that is in process of being acculturated to the West."⁵⁴ More recently, Paul Feyerabend proposes that mentalistic language be abandoned entirely, at least for scientific and philosophical purposes. He allows that a theorist might *redefine* mentalistic expressions in terms of brain states, "if he intends to perpetuate ancient terminology,"⁵⁵ although Feyerabend does not seem to think there is any advantage to be gained from this.

There are many difficulties with such views, the most obvious one being that it will be a very long time before we know enough about the brain to begin to develop such a language and that such a language, if ever developed, would be of a complexity far beyond the intellectual powers of any

but the most brilliant neurophysiologists. But even if we could manage such a language, it would not do the many things that our mentalistic language does. For example, when I report that I suddenly remembered that Henry was sick, the intentionality of this report, i.e., that it is about Henry and his sickness, is an essential part of it. This intentional feature is lost if we simply report that a particular neural event had suddenly occurred; such a report would not be about Henry at all, only about a brain event. Of course we could always give these new functions to brain-event reports, but that would be to redefine physicalistic expressions, instead of redefining mentalistic expressions, leaving us where we began.

J. J. C. Smart takes a more tolerant view of our ordinary mentalistic language.⁵⁶ He takes it to be a way of talking about brain events, but an inexact, indefinite, vague way. When we report, say, a pain, instead of reporting that a particular brain event occurred we report (for lack of better information) that something happened such as happens when, for example, a pin is stuck into us. I suppose that Smart might analyze my report that I suddenly remembered that Henry was sick as follows: something went on which is like what goes on when someone says to me, "Let me remind you that Henry is sick." In such cases we indicate the event by pointing to the circumstances which typically tend to cause it; we might also point to typical effects.

The rejoinder to this must be blunt.⁵⁶ Such is *not* what we mean when we use such language. If it were, then the congenitally blind neurophysiologist would understand the report of a red image as well as and in exactly the same way as the rest of us do. But this is simply not the case. He may know far better than I what goes on chemically and electrophysically in the brain when I have that red image, but he cannot understand in the way I can what it is for something to *be a red image*.

I suspect that the reply of Identity Theorists would be equally blunt. So much the worse for ordinary mentalistic expressions. They would advise ordinary language to do what Broad advised common sense to do, "to follow the

⁵⁴ B. A. Farrell, "Experience," *Mind*, vol. 59 (1950); see p. 195.

⁵⁵ Paul K. Feyerabend, "Comment: Mental Events and the Brain," *The Journal of Philosophy*, vol. 60 (1963); see p. 296.

⁵⁶ J. J. C. Smart, "Materialism," *The Journal of Philosophy*, vol. 60 (1963), pp. 654-655, 661-662. In addition to Smart's publications already mentioned, see "Further Remarks on Sensations and Brain Processes," *The Philosophical Review*, vol. 70 (1961), pp. 406-407.

⁵⁷ For more extended argument on this point, see my article, "Mental Events and the Brain," *The Journal of Philosophy*, vol. 60 (1963).

example of Judas Iscariot, and 'go out and hang itself';"⁵⁶ if it were not reformable in the physicalistic way they suggest. Even Smart admits that "we cannot . . . hope . . . to reconcile *all* of ordinary language with a materialist metaphysics,"⁵⁷ any more than Newtonian theory can be reconciled with general relativity theory.

This last move brings us to the end of this particular road. For the only reply can be that to abandon mentalistic expressions is to render us incapable of talking about events which clearly and undeniably occur, viz., mental events. That is just what Materialists deny. It is worth pointing out, in the presence of this impasse, that the Materialist has here abandoned the Identity Theory. Instead of holding that mentalistic expressions do have reference (albeit physical reference), he now denies that they have reference at all. So, just as Feigl's idealistic version of the Identity Theory ends up being an Interactionist Theory, so this version ends up being out-and-out Materialism and not an Identity Theory at all.

A Neutral Version of the Identity Theory

We are trying to understand the relation between mentalistic expressions and physicalistic expressions, assuming the referent of each is the same set of events. A popular view is that mentalistic and physicalistic expressions describe different "aspects" of one and the same event, or describe it "from different points of view" or "at different levels." The force of such suggestions is that the difference between the mental and physical is not an intrinsic difference, not a difference in the nature of things, not a difference in features or properties of things, but a difference relative to human purposes, outlook, conceptual scheme, or frame of reference. But the trouble with such approaches is that they are *deplorably vague*. Suppose we do say that we can look at a particular event as a physical occurrence in a physical brain or we can look at the same event as a mental occurrence in a mind. How does such a remark make anything clearer? How are these two "ways of looking" related?

One suggestion frequently made is that in one case we look at the event "from the outside" and in the other "from the inside." R. J. Hirst puts it this way:

The various experiences of imagining, remembering and thinking . . . are the inner aspect of these various episodes and activities in a person's life and are what the person concerned experiences as an actor and not spectator in these situations, whereas brain activity is an outer aspect and is what can be scientifically observed or inferred by others.⁵⁸

The question is what "inner" and "outer" mean here. Taken literally they suggest the contrast of a man free to roam about inside his house as opposed to the man who can only peer in from the outside. But this utterly misrepresents the state of affairs. My "inner" aspect, in this sense, is much more accessible to the surgeon than it is to me; he has the X rays, probes, electrodes, saws, scalpels, and shears for getting at my insides. Presumably, the contrast which is meant is something like the "awareness" of brain events as opposed to the events themselves. But then we are still left with the fundamental problem of the relation between events of consciousness or awareness (i.e., mental events) and brain events.

Some prefer to resort to machine analogies here. For example, mental events are to be compared to occurrences in an internal scanning device or in an internal register. But this will not do, because such occurrences in the scanner or register would be in principle observable by an external observer whereas my mental occurrences are not so observable. There is also Putnam's comparison, already discussed, of mental events with the "logical" states of a machine.

Here we have the basic difficulty in the various "double-aspect," "double-viewpoint," or "double-language" versions. None of them makes clear how this doubleness is to be understood. When we say that the Morning Star and the Evening Star are one and the same star, we can explain how the two referring expressions are related. But no Identity Theorist, to my knowledge, has done this for mental and physical expressions.

The Location Problem

So far we have been considering the problems which arise for the various formulations of the Identity Theory. The time has come to attack the theory directly, by bringing up the major difficulty in any theory of this sort, the location prob-

⁵⁶ C. D. Broad, *The Mind and Its Place in Nature* (London, 1925), p. 186.

⁵⁷ "Materialism," *op. cit.*, p. 661.

⁵⁸ R. J. Hirst, *The Problems of Perception* (London, 1959), p. 195.

lem.⁵⁹ The physical events which are intimately connected with my having particular mental events have some definite location, probably in the brain. This is not to say that they are localized in some small part of the brain or even in a number of small parts of the brain. Perhaps they are spread throughout large parts of the brain; perhaps they are fields; perhaps they include the nonoccurrence of certain events. Nevertheless such phenomena occur not only at some time but also in some place or places in the brain. However, so far as thoughts are concerned, it makes no sense to talk about a thought's being located in some place or places in the body. If I report having suddenly thought something, the question *where* in my body that thought occurred would be utterly senseless. It would be as absurd to wonder whether that thought had occurred in my foot, throat, or earlobe as it would be to wonder whether that thought might have been cubical or a micron in diameter.

Since brain events occur somewhere in the brain but mental events do not, it follows that they cannot be identical. Two things can be (unproblematically) identical only if, among other things, wherever one is the other is. If it were true that the president of the United States were in Washington but it was not true that the commander-in-chief of the Armed Forces was in Washington, then one could not say that the two were identical.

We do give location to bodily sensations. But for the most part we do not locate them in the brain. So here again, except in the special cases of some headaches, there should be no temptation to identify them with brain events.

It might be argued that identity of location is not a necessary condition. Don F. Gustafson claims that if the assertion that a person had a mental event of a certain sort turns out to be true if and only if the assertion that he had some physical event of a certain sort is true, then we would say that the two assertions assert one and the same fact.⁶⁰ But I do not see that this material equivalence is sufficient for identity, nor even nomological equivalence. Imagine a case where some underlying condition, say a particular gland defect, always produces two different phenomena, say a particular sort of skin eruption and a particular

sort of blood condition, and these phenomena are produced by that condition alone. Then a person would have the skin eruption if and only if he had the blood condition. But we would obviously be dealing with different facts (and different events). Other sorts of examples could easily be given. So material or nomological equivalences are not enough. Same *place* (and *time* too, of course) would also be necessary for identity.

Another argument might be offered to show that identity of location is not a necessary condition. Such things as temperature, potential energy, solubility, electrical resistance, and viscosity do not have a location,⁶¹ yet they may still be identical with micro-states of a physical object. Since we are considering mental *events*, let us consider the comparable events, *changes* in temperature, potential energy, solubility, etc. Surely we can ask *where* in the object these changes occurred. Certainly if we identify temperature with some property of the molecules, perhaps their average kinetic energy, then the change in temperature must occur where the change in the molecules occurs. If we are not willing to locate the change in temperature there, we must not say it is identical with the change that does occur there. To take another example, when a person goes into shock, we do not locate the shock in the vascular system (nor anywhere else in the body). That shows it is wrong to identify shock with, say, a collapse in blood pressure; the latter is, at best, the *cause* of shock.

It has often been suggested to me in discussion that location ceases to be a necessary condition if we talk not about events but about *states*, since states do not have location (although the things they are states *of* may have location). I have two objections to this move. First, it seems to me that the sensations, feelings, and thoughts we have been talking about are more naturally taken to be events rather than states. Second, if we take them to be states, in order to show identity we must show that they are states *of the same thing*. And here a similar conceptual point arises, namely that it does not at present make sense to ask what part of the body a thought is a state *of*. It would be as nonsensical to ask if a thought is a state of the head as it would be to ask if it occurred in the head.

⁵⁹ I have argued this point at length in my article, "Could Mental States Be Brain Processes?" *The Journal of Philosophy*, vol. 58 (1961), pp. 813-822.

⁶⁰ Don F. Gustafson, "On the Identity Theory," *Analysis*, vol. 24 (1963), pp. 30-32.

⁶¹ This is pointed out by Ernest Nagel in "Are Naturalists Materialists?" *The Journal of Philosophy*, vol. 42 (1945), reprinted in Nagel's *Logic Without Metaphysics* (New York, 1956). Nagel uses this point only to show that properties and processes may belong to a physical object even if they lack spatial dimensions.

Because of this non-locatability of mental events as we presently conceive them, we cannot identify mental events and brain events. But our concepts may change, perhaps are changing already. We might adopt a convention for locating mental events in space, perhaps for locating them in that part of the brain where the associated brain events occur.⁶² Such may gradually come to pass. Our concepts do change, after all. Talking, as we do now, about machines which learn, play checkers, compose music, etc., involved a shift in language. Yet these shifts were perfectly sensible. The same may happen with mental events. We may come to talk about them as occurring in the brain if developments in neurophysiology make such a shift desirable. Such a change in our concepts would make the Identity Theory quite plausible. For it would involve thinking of events like thoughts in a more physical way, more like public happenings occurring in a particular place (and perhaps having shape and size, even color, texture, etc.). Thus it would be less of a wrench to our conceptual scheme to think of them as identical with brain events.

Even if we modified our concept of mental event, the peculiar features of mental events would remain, however, especially what was referred to in Section I as the "asymmetry" of mental reports, namely that I can know without observation that I am having a mental event. If we begin to think of these events as physical, then it will turn out that I can have *privileged access to certain physical events*. In other words, as we begin to think of mental events in a more physical way we at the same time begin to think of (certain) physical events in a more mentalistic way. As we drift toward Materialism, Materialism must drift toward Idealism.

But this is to anticipate the future. As things now stand, the Identity cannot be maintained, since our language and conceptual scheme does not allow us to locate mental events in the brain. So an Identity Theory in any of its variants cannot be accepted.

IV. CAUSAL THEORIES

The Theories

In this section we shall consider whether there is a causal connection of some sort between mental events and brain events. From the start we labor

under the handicap of not having a very good understanding of the difficult concept of *cause*. A clear account of that notion would make our task much easier, but I do not know of one. Very roughly, I understand by "cause" of something some prior event or state which, given the circumstances, is necessary and sufficient for the existence of the thing. If, given the circumstances, *A* and *B* are constantly correlated and if when we keep those circumstances constant, the occurrence of *A* is followed by the occurrence of *B* and the non-occurrence of *A* by the nonoccurrence of *B*, then we have good *evidence* that *A* is the cause of *B*.

When we try to determine if mental events and brain events are causally connected we find ourselves immediately frustrated by the lack of relevant factual information. We do not know if there is any kind of constant, universal correlation between mental events and brain events. We do not even know very much about what *kinds* of brain events *might* be related to mental events. Assuming we decide upon the relevant brain events, will the correlations with mental events be one-to-one or many-to-one or even many-to-many? How about timing? Will brain events and their presumed mental correlates be simultaneous or will there be temporal priority? Considering brain events by themselves, will we discover laws which govern their succession? Laws governing the succession of mental events? We can hardly begin to make informed guesses about such questions, and yet it is obvious that the answers to such questions are absolutely necessary to coming to any definite conclusions about the relation of brain and mental events. At the very best, at present, we are dealing with what are at present highly speculative issues. Even here, however, it is useful to apply logical considerations to determine if the issues are in any way resolvable, what sort of new data might make one speculation more likely than another, whether there are general objections to any of them, and whether any conflict with what we already know. It is these logical issues which we shall pursue.

Among the dualistic (i.e., nonidentity) theories, the ones seriously discussed today are Parallelism, Epiphenomenalism, and Interactionism. Parallelism holds that every mental event is correlated with some one brain event (or a disjunction of them) in such a way that whenever that mental

⁶² My claim ("Could Mental States be Brain Processes?" *op. cit.*, pp. 816-818) that such a convention could be adopted is contested by Robert C. Coburn, "Shaffer on the Identity of Mental States and Brain Processes," *The Journal of Philosophy*, vol. 60 (1963). I reply in "Mental Events and the Brain," *op. cit.*, p. 165.

event occurs that brain event (or one of the disjunction) occurs, but *there is no causal connection* between them; they just happen to occur together. Epiphenomenalism agrees with Parallelism that there is the correlation, but adds that it consists of a one-way causal connection in which every mental event is caused by some brain event and no brain event is in any way causally affected by any mental event. Interactionism holds that there are causal connections both ways so that some mental events are caused by brain events and some brain events are caused by mental events.

Many philosophers have held that there is no way of deciding between these positions. However, as we shall see, there are relevant considerations which might help us to decide.

Parallelism Rejected

In recent years Parallelism has tended to be eliminated from discussion. The kind of evidence there would be for it, constant correlation, would in any other context be taken to indicate a causal connection of some sort. The situation is not at all analogous to that of Leibniz' two perfect clocks which are forever in phase. In the case of the clocks we can appeal to the mechanisms of the two clocks to *explain* the correlations of their chimings. But it is an empirical fact that minds do not form a self-contained isolated system such that every mental event can be explained merely by prior states of mind. So the constant correlation would have to be sheer accident. Some philosophers have been driven to this desperate conclusion because they believed that mental events and brain events were too dissimilar for there to be a causal connection. But as Martha Kneale points out, we are not entitled to have *a priori* convictions about what events can and cannot enter into causal relations: "as empiricists, we should be prepared to find them anywhere."⁶³ To be sure, such a causal relation would be different from any presently recognized ones. But that it is radically different is not sufficient reason to refuse to call it causal at all.

It is true that constant correlation would not *entail* that there is a direct causal connection, i.e., that one is the cause of the other. Early symptoms of a disease are not the cause of the later developments even though they may always precede those later developments. But they are causally connected in that they earlier and later effects of an underlying pathological condition. So one might

argue from the constant correlation of mental and brain events that they are indirectly causally connected by their relation to some underlying thing. Since such a third thing, neither mental nor physical, but affecting both, is unlikely, it becomes reasonable to think there is a direct causal connection between them.

Interactionism vs. Epiphenomenalism

In deciding between Epiphenomenalism and Interactionism, certain empirical information would be highly relevant. For example, suppose it turned out that there existed a slight time gap such that a mental event always came *after* the brain event with which it was constantly correlated. This would indicate that it was the brain event which caused the mental event, i.e., Epiphenomenalism. However, it is unlikely that we shall ever be able to so refine our experimental techniques that we shall be able to distinguish simultaneity from slight priority. A more interesting possibility is that it would turn out that certain brain events are not correlated with any particular prior brain events but are correlated with certain prior mental events. It is conceivable that the following situation might occur: a difficult question occurs to me (and my brain undergoes an event which in all humans is correlated with that question occurring) and then I come up with an insightful answer (and my brain undergoes an event which in all humans is always correlated with that answer occurring) but there is no causal connection, either direct or through intermediaries, between the first and the later brain events. If such a sequence occurred, the most natural interpretation would be Interactionist, namely that here the brain idled, as it were, while the insightful (mental-event) thought of the solution occurred, and this thought then produced the correlated brain event. Of course it would be possible for the Epiphenomenalist to hold out for a long time, claiming that we had not looked long enough for intervening brain events which causally connected the first and later brain states. He might even claim that there was real *indeterminism* in the brain. However, there could come a point when this would be sheer dogmatism.

If there are these gaps in the sequence of brain events which are inexplicable physiologically but explainable on an Interactionist hypothesis, it will be a very long time indeed before neurophysiology can be developed to the point where it

⁶³ Martha Kneale, "What Is the Mind-Body Problem?" *Proceedings of the Aristotelian Society*, vol. 50 (1949-1950), p. 116.

will be able to determine this with any confidence. Presumably it is the hope of neurophysiologists that no such cases will arise. They hope that a complete neurophysiological account can be given of brain phenomena, so that everything which happens in the brain would be causally explainable in terms of prior brain events and physical inputs into the brain. What would be the implications if such turned out to be the case?

Given this completeness of the neurological story, plus our earlier assumption that some brain events stand in constant correlation with certain mental events, Epiphenomenalism would clearly be possible. But Interactionism is not as yet ruled out. The Interactionist could maintain that in some cases it is really the mental events which do the work and the brain events correlated with them are merely their effects. As it stands, the issue cannot be decided because the correlations we postulate do not settle the question of which caused what.⁶⁴ The issue is real enough, however difficult it seems to settle it. The Epiphenomenalist, in claiming that the brain is the basic cause, says that human behavior and the whole of human history would be in no further way changed if there never had been a mental event. The Interactionist denies this, saying that since mental events do have effects their absence would have made a great difference. But since there is no conceivable way of performing the crucial experiment of observing what happens to the brain when we prevent mental events from occurring, there is no way of deciding this controversy directly.

All is not lost, however. The following *indirect* consideration would seem to favor Epiphenomenalism. It is reasonable to believe that there do exist genuine causal sequences of brain events, and that we could have evidence for the existence of such causal sequences of brain events. The Interactionist would have us believe of *certain* sequences of brain events, for which there would be exactly the same sort of evidence which is usually taken to indicate causal connection, that these particular sequences were not really causally connected at all. They would have us believe that in just this special case the brain events were effects of some quite different sort of causation. And they could give us no evidence for this astonishing hypothesis. Surely it would be good scientific sense to treat the brain events which are correlated with mental events in the same way as other brain events unless we had

special evidence that they were different. Of course if they were different, if, for example, they did *not* stand in a lawful relation with their predecessors, if there were inexplicable gaps in the brain sequences, then Epiphenomenalism would begin to look dubious.

The Paradox of Epiphenomenalism

Many philosophers find Epiphenomenalism too paradoxical to take seriously. On that hypothesis, even if there had never been a single mental event, the whole of human history would have been the same, there still would have been developed a language with expressions for reporting mental events, and, incredibly, there would even have been those verbal interchanges which constitute our arguments about the very existence of mental events. To suggest otherwise would be to admit that the existence of mental events makes a difference, and no Epiphenomenalist will admit that. Yet how could there be no mental events and yet, to take the most incredible element in this story, men assert that they themselves certainly had them, wonder whether their fellow men had them, and argue about the issue?

This paradox has two components, a conceptual one and a factual one. The conceptual one is more easily resolved. It concerns the question whether it would be possible for there to be a world just like ours except that it contained no mental events and the connected question how we should *describe* such a world. Let us recall the machine of Section II above which is so built that it can use mentalistic language. The question whether such a machine actually had mental events becomes the question whether the machine which used mentalistic language would *understand* that language in the way we do. The same principle is involved in the question whether the world could be the same except that there were no mental events. If there were no mental events but everything else went on exactly the same, then certain utterances used on certain occasions would not have the same meaning they do for us. For example, if a man uttered the expression, "It just occurred to me that . . .," it might have some purely behavioral significance or might simply be a sign (like a man's suddenly bringing his hand to his forehead). But such expressions might still be used. So in terms of overt and interior behavior everything might be exactly the same. Wittgenstein gives the following example:

⁶⁴ This has most recently been argued by John Lachs in "Epiphenomenalism and the Notion of Cause," *The Journal of Philosophy*, vol. 60 (1963), pp. 141-146.

We say: "the cock calls the hen by crowing"—but doesn't a comparison with our language lie at the bottom of this?—Isn't the aspect quite altered if we imagine the crowing to set the hens in motion by some kind of physical causation?

But if it were shown how the words "Come to me" act on a person addressed, so that finally, given certain conditions, the muscles of his legs are innervated, and so on—should we feel that the sentence lost the character of a *sentence*?⁶⁵

So we can imagine things going on as a mere physical interaction between bodies. Yet everything would be quite different in that our descriptions, in our language *as we now understand it*, would be quite different.

There is still a factual question which remains. One might argue that it is just *false* that mental events are never causes. Are we not perfectly familiar with countless indisputable cases in which it is appropriate to say that mental events cause other mental events or cause physical events? A mother thinks of her child's not being home yet and that thought causes a chill to run down her spine or causes her to feel a pang of fear. A man is at the dentist's and a sudden pain causes him to wince. A woman's anger may cause the blood to run to her cheeks or cause her lower lip to tremble. The sudden realization of his peril may cause a man to stop dead in his tracks or take to his heels. Here we seem to have a number of perfectly ordinary cases of mental events which are causes, sometimes of other mental events and sometimes of physical events. How can the Epiphenomenalist claim that we are *always mistaken* when we cite a mental event as a cause? That he does claim this is a "great paradox of Epiphenomenalism."⁶⁶

Insofar as Epiphenomenalism asserts that mental events are never causes, it does seem to be in flat contradiction to our ordinary descriptions of familiar experiences which allow us to say, in certain circumstances, that mental events are sometimes causes. It looks as though we must either reject Epiphenomenalism or else say that it has never been the case that a sudden pain has caused a person to wince. Some philosophers have boldly accepted the latter alternative. Broad points out that we must distinguish between what he calls "*de facto* invariable accompaniment" and "causal

conditioning."⁶⁷ The invariable accompaniment which we have in the case of putative mental causation produces the *illusion* of causal connection.

Many philosophers would reject this distinction. Thus Feigl says, "For empiricists holding an essentially Humean conception of causality, it is then quite permissible in this sense to speak of the causal efficacy of mental states."⁶⁸ For my own part, I believe the distinction can be made, for I should wish to insist that some events which invariably precede others are *not* causes of those things; I have in mind such cases as certain signs or symptoms of occurrences, Leibniz' two clocks, successive manifestations of an underlying process, etc. So Broad's distinction is acceptable. However, when we have invariable accompaniment it is natural and reasonable to assume that there is direct causal connection unless we have special reasons for accepting one of the more complicated models. The demonstration that sequences of brain events are fully explicable in terms of physiological laws might provide such special reasons. But we are still a long way from that. To insist at this time that putative mental causes are merely invariable accompaniments but not causes seems arbitrary.

Let us note that to accept it as a fact that mental events are causes is not quite to embrace Interactionism (assuming we believe that sometimes brain events cause mental events). For to say that a mental event causes the wince is not to imply logically that a mental event causes a *brain event*. Still it would be easy enough to convince someone of this by tracing back the wince to the brain. So while our ordinary ways of talking are not *technically* Interactionist, still they commit us to Interactionism when we add the known facts of physiology which link bodily changes with the brain. It is important to note this because it shows that the paradox of Epiphenomenalism arises only if we graft onto our ordinary language certain technical physiological facts, and this might be unfairly to wrench ordinary language from its proper context. Yet it does seem that it is only proper to take account of such new discoveries. If they show ordinary language to have certain consequences, then we could accept these consequences or else reform or reinterpret ordinary language.

⁶⁵ L. Wittgenstein, *Philosophical Investigations* (Oxford, 1953), par. 493.

⁶⁶ William C. Kneale, "Mental Events and Epiphenomenalism," in *The Philosophy of C. D. Broad*, ed. Paul A. Schilpp (New York, 1959), p. 453.

⁶⁷ C. D. Broad, "A Reply to My Critics," in *The Philosophy of C. D. Broad*, *op. cit.*, pp. 791-794.

⁶⁸ Herbert Feigl, "Mind-Body, Not a Pseudoproblem," in *Dimensions of Mind*, *op. cit.*, pp. 36-37.

"Volitions"

In this paper I have taken thoughts and feelings as my cases of familiar mental events, but I have ignored that third member in the hallowed trio of "cognitions, affects, and volitions." Yet many philosophers have taken volition as the very paradigm of mental causation, so a word should be said on this topic, at least to justify its omission.

First it is clear that volitions and related things like motives, intentions, and reasons are not necessarily or always *mental events*, in the sense here intended of datable occurrences which the subject knows immediately and incorrigibly. In fact it is likely that a "volition" is nothing more than a fiction.⁶⁹ There are mental events which might serve, as when we think: "I must have that," or "No more for me," or "Time to get up," and then do the appropriate action, but even when we have such thoughts they are often utterly idle thoughts and have no connection with action, and few of the willful and voluntary things we do are preceded by such thoughts.

Second, even if we admit the existence of volitions as mental events, it has been objected by many that a human action is quite different from a physical occurrence.⁷⁰ For one thing, an action is something done by a *person* whereas the physical occurrence, a bodily movement or change, is always something which is done by or which happens to some part of the body (e.g., his finger moved, his heart beat faster). Nor does an action entail some particular bodily movement, since for any action involving bodily movement there are always a number of different movements by which one could perform the action.⁷¹ So even if it could be shown that an *action* had a mental cause it would not follow from this alone that a bodily change had a mental cause.

Finally, it is a matter of much controversy today whether the relation between volition, motives,

intentions, and reasons, on the one hand, and overt action on the other is to be construed as a *causal* relation. Some argue that motives, intentions, and reasons cannot be construed as causes,⁷² others argue that volitions cannot be causes,⁷³ and others argue that actions done for a reason cannot have causes.⁷⁴ With all of these very thorny knots to unravel, the case of volitions is not a very good one to take as a paradigm of interaction between the mental and the physical.

Can Epiphenomenalism and Interaction Be Reconciled?

Epiphenomenalism says that mental events are never causes. Familiar experience seems to show us countless cases of mental events as causes. Must we reject one or the other? One way out of this dilemma would be to show that there is no contradiction because we have here two different concepts of cause: mental events are not causes₁ but they are causes₂. What could be the difference? Physical causation vs. non-physical causation?⁷⁵ But the Epiphenomenalist insists that brain states can cause mental states, so it must be a kind of causation between physical and mental. Why should it be causation in a different *sense* if it goes in one direction rather than the other? Could it be causation *à la* general laws vs. causation without laws?⁷⁶ There is no reason to think there are laws in one case but not in the other. How about this: the brain state is a cause in the sense that it is (presumably) an entirely sufficient condition for its correlated mental state whereas in the examples cited above the mental state is only a necessary condition. But there is still a contradiction because something cannot be a sufficient condition for a thing if something quite different is a necessary condition for the same thing. Suppose we appeal to the type of cause called by G. E. M. Anscombe a "mental cause."⁷⁷ An example of such causation would be "The thought of his predicament caused

⁶⁹ Cf. G. Ryle's section on "The Myth of Volitions" in *The Concept of Mind*. A useful but admittedly artificial word, "volit," is introduced by Anthony Kenny in his excellent *Action, Emotion and Will* (London, 1963), pp. 214-215.

⁷⁰ D. W. Hamlyn, "Behavior," *Philosophy*, vol. 28 (1953), pp. 132-145. See also R. S. Peters, *The Concept of Motivation* (London, 1958), pp. 12-14, and A. I. Melden, *Free Action* (London, 1961).

⁷¹ Peters, *op. cit.*, pp. 12-13.

⁷² Cf. Philippa Foot, "Free Will as Involving Determinism," *The Philosophical Review*, vol. 66 (1957); G. E. M. Anscombe, *Intention*, *op. cit.*

⁷³ e.g., Melden, *Free Action*, *op. cit.*

⁷⁴ Peters, *Concept of Motivation*, *op. cit.*; Anscombe, *Intention*, *op. cit.*

⁷⁵ H. L. A. Hart and A. M. Honore, *Causation in the Law* (Oxford, 1960), pp. 48-57.

⁷⁶ William Dray, *Laws and Explanations in History* (Oxford, 1957).

⁷⁷ G. E. M. Anscombe, *Intention*, *op. cit.*, pp. 16-18. For further discussion of "mental cause," see Jenny Teichmann, "Mental Cause and Effect," *Mind*, vol. 70 (1961), pp. 36-52, and D. F. Pears, "Causes and Objects of Some Feelings and Psychological Reactions," *Ratio*, vol. 4 (1962), pp. 91-112.

me to laugh." Here the person who makes the report knows non-inferentially and incorrigibly what caused him to laugh. Even if we allow that this represents a special kind of causality rather than merely a special way of knowing or more familiar and easily known cases, it still does not help us because it cuts across the distinction between causes₁ and causes₂. The thought which caused my depressed feeling may not be a "mental cause"; the physical event I perceive may be a "mental cause," as in Anscombe's case, "The leap and loud bark of that crocodile made me jump."⁷⁸ Although I do not know of any other promising ways of distinguishing between cause₁ and cause₂, it still remains a possibility that someone will find one. Until that time I think we are entitled to say that the sense in which the Epiphenomenalist talks of mental events as *caused* involves the same sense of "cause" as that in which the ordinary man talks of mental events as causes. So we are still left with our paradox.

Because of our present scientific ignorance, we are really in no position, yet, to decide between Epiphenomenalism and Interactionism. If there turn out to be gaps in the physiological account which can be explained only by appeal to mental events, then Interactionism will be reasonable. But if the physiological account is completable, then both Epiphenomenalism and Interactionism would fit the facts. Epiphenomenalism would have in its favor the conceptual economy of having one kind of explanation, a physiological one, for anything which happened in the brain. Interactionism would have in its favor its consistency with ordinary ways of describing familiar phenomena. I do not see that either alternative is clearly superior.

As one final attempt to make peace between these two theories, let me suggest the following. Interactionism claims that for certain physical occurrences, mental events are sufficient causal conditions. Epiphenomenalism claims that for those physical occurrences, physical events are sufficient causal conditions. So far there is no conflict, since there is no reason why there might not be *two* sets of sufficient conditions.⁷⁹ Imagine a car on a hill, held in place by its brakes, by blocks jamming the wheels, and by its engaged clutch. Supposing that each might be sufficient to keep the car from rolling, the question of which of them

really keeps it from rolling would be silly. Similarly mental events and their concomitant brain events might *each* be sufficient causes. The conflict between the theories arises because each theory makes a further claim, namely, that its putative cause is a *necessary* condition as well, so that if it had not occurred then no matter what else happened the effect would not have occurred. And this claim is always contrary to fact by virtue of the assumed constant correlation of brain event with mental event. So we could never test what would happen if we took one away and produced the other. Since this further claim is in fact untestable, it makes it impossible in fact to decide between Interactionism and Epiphenomenalism. If we drop this further claim, *both* theories become acceptable. We might call their joint assertion the Dual Causation Theory, the theory that certain brain occurrences are caused by *both* mental events *and* brain events, each one of which alone could have caused it. Of course, the Dual Causation Theory could no more be established than either Interactionism or Epiphenomenalism. But it would account for familiar mental causation and also the completeness of physiological causation, if future evidence point to the latter.

Summary

Among the mind-body theories which take mental events as non-identical with brain events, the main contenders are Epiphenomenalism and Interactionism. If future information about the brain indicates that there are sequences of brain events which are inexplicable physiologically but explicable in terms of prior *mental* events, then Interactionism will be the more plausible view. If it seems likely that all brain events are explicable physiologically, then either view is still possible. We could hold that mental events are inefficacious by-products of physiological occurrences or we could hold that they are indispensable causal intermediaries. Epiphenomenalism would have the advantage of offering a simpler account, explaining the physiological only in physiological terms. But it still would be in conflict with our ordinary descriptions of everyday experience, so if we accepted Epiphenomenalism we would have to abandon or reinterpret our ordinary ways of speaking. The alternative would be to continue to

⁷⁸ *Op. cit.*, p. 15.

⁷⁹ See the Section "Of Plurality of Causes" in J. S. Mill, *A System of Logic*, book III, chap. X. Such cases are discussed briefly by Michael Scriven in *Philosophy and History*, ed. Sidney Hook, pp. 348-349.

II. MEDITATIONS LEIBNIZIENNES*

WILFRID SELLARS

I

MY purpose in this paper is to explore the thesis, so central to Leibniz' philosophy, that the world in which we live is but one of many possible worlds, decidedly more numerous than blackberries. The exploration I have in mind is partly historical, concerned with the questions: "How exactly is Leibniz' thesis to be understood?" and "How did he defend it?" But I also have in mind the question, "Is this thesis, or something reasonably like it, true?" My starting point will be Leibniz' contention—so brusquely rejected by Arnauld in his first letter—that "the individual concept of each person includes once for all everything which can ever happen to him."

Now the phrase "individual concept" will be at the center of the stage, once the argument is fully under way. For the moment, it will suffice to characterize the individual concept of a substance as the sense of God's proper name for that individual; thus, the individual concept of Julius Caesar is the sense of the divine name for the individual substance we refer to as Julius Caesar. Although Leibniz insists that we have a confused grasp of this sense, which consists of our *petites perceptions* in so far as they represent Julius Caesar, this confused grasp of the individual concept is not, of course, *our* concept of Julius Caesar. For the sense of the term "Julius Caesar," as we use it, is not, strictly speaking, an individual concept at all but—one is tempted to say—a peculiar kind of general concept which applies to many *possible* individuals, though only to one *actual* individual. And, indeed, one is tempted to say that for Leibniz, the "names" we use are not really names at all, but a peculiar kind of general term.

But more about names, divine and human, individual concepts, and possible individuals

later. For our present purposes, the important thing about the individual concept of an existing substance is that though as *concept* it exists in the *divine understanding*, it exists *in re* as the *nature* of the substance. This gives us a second formulation of the contention which so startled Dr. Arnauld, to wit, "*the nature of each individual substance includes once for all everything which can ever happen to it.*"

The notion of the nature of an individual substance is a venerable one, though not without its puzzles. But this notion has obviously taken a new twist in Leibniz' hands. Leibniz was not the first to conceive of the *nature* of an individual substance as accounting for its individuality. He was, however, the first to see clearly that the individuality of a substance can only be understood in terms of *episodes* in its history, and to conclude that if the nature of a substance is to account for its individuality, it must account for episodes, and not merely the capacities, powers, dispositions—all, in principle, repeatable—which were traditionally connected with the natures of things.

If we meant by the *nature* of an individual, the criteria in terms of which we identify it as that individual, there would be no puzzle to the idea that natures individuate. But, of course, this is not how we use that expression. We may identify a certain automobile as the one owned by Smith, but in no ordinary sense of the term is *to be owned by Smith* the nature of the automobile. However difficult it may be to make the notion of the nature of a thing precise, this nature is not that in terms of which we identify it, but that in terms of which, *if we but knew it*, we could explain why it behaves as it does in the circumstances in which it is placed.

Now if we take as our model for interpreting the physical conception of the nature of the substance the kind of account we find in Broad,¹ we can by suitable oversimplification construe the nature of a

* A shortened form of this essay was read as the opening paper in a symposium on Rationalism at the May, 1958, meeting of the American Philosophical Association. In preparing the manuscript for publication I have limited myself to stylistic changes and, where matters of substance were involved, to the omission of less fortunate passages. I have, however, added a brief discussion (Section III) of Leibniz' general theory of relations to provide a background for the more specific discussion of causality.

¹ C. D. Broad, *Examination of McTaggart's Philosophy*, vol. I (Cambridge, 1933), pp. 264–278.

substance *S* in terms of facts of the form "if at any time *S* were to be involved in an episode of kind *E*₁, it would be involved in an episode of kind *E*₂." Thus, suppose that on a certain occasion *S* is found to have been involved in an episode kind *E*₂, then the nature of *S* would account for this fact in the sense that if we knew the nature of *S* and if we were to discover that *S* had obviously been involved in an episode of kind *E*₁, we would be in a position to say

S was involved in an *E*₂ because it was previously involved in an *E*₁

If one accepts the above as a crude model for the classical account of the nature of the thing, the first thing one is tempted to say about Leibniz' conception of the nature of a substance is that it not only provides the general hypotheticals, but the episodic premisses as well. Such a nature would, in Hegelian terms, be a set of syllogisms *in re*.

II

How are we to account for this strange twist, in Leibniz' hands, to a familiar notion. It might be thought that the explanation is to be found in his denial of interaction. If the explanation of what goes on in a substance is not to be found in what goes on in another substance, must it not be found in that substance itself, and hence in its nature? Now there is indeed a connection between his conception of the nature of the substance and his denial of the reality of relations between substances. But the latter by itself does not account for the peculiarity of his view as can be seen when we notice that Spinoza is in his own way committed to the idea that the nature of a substance provides not only the hypotheticals (laws) but *affirms the antecedents* as well. For, conceived under the attribute of extension, the nature of Spinoza's one substance specifies not merely that if the physical world were at any time to be in a certain state, it would subsequently be in a certain other state, but specifies whether or not, at a given time, it is in the former state. The nature of substance not only provides the *if* but turns it into *since*.² At the heart, then, of Spinoza's conception of the nature of substance is the demand that the occurrence of any *episode* has (in principle) an explanation which is not simply of the form

This episode because that episode

Such an explanation is, of course, legitimate as far as it goes. It is, however, a *relative* explanation of one episode in terms of another. Spinoza demands, in Kantian terms, that the series of other-grounded episodes must have its ground in something, obviously not an episode, which accounts for its own existence. This self-explainer is substance (*Deus sive Natura*); and, of course, in thinking of it as a self-explainer, he is thinking of an argument *in re* of which one premiss says that substance *can* exist, another premiss says that what *can* exist *must* exist if there could be nothing incompatible with its existence, and another premiss, itself a conclusion from a prior argument, says that nothing could be incompatible with the existence of substance. Let us be quite clear that whatever rationalists may have said about abstractions, they were precluded from holding that the *esse* of possibilities is *concupi*. Now Leibniz makes exactly the same demand with exactly the same result. Reality provides the principle and affirms the antecedent of an argument *in re* which proves the existence of any episode which belongs to the history of the actual world. But, unlike Spinoza, he offers a complicated story which makes some sense of the idea that this might be the way things are—whereas Spinoza ultimately rests in the assurance that it *must be so* if the world is to be intelligible.

But, of course, the idea that the actual course of events in the world is the only possible course of events is *prima facie* so absurd that the principle of sufficient reason on which it rests would have no plausibility at all unless some meaning could be given to the idea that other courses of events are possible—even if *in the last analysis* they aren't really possible. Leibniz offers such an account.

A useful way to get the hang of this account is to conceive of a philosopher who is a blend of Leibniz and Spinoza; let us call him Leibnoza. Leibnoza, unlike Leibniz, is happy about the interaction of finite substances. He conceives of the universe as a set of interacting substances whose natures are *hypotheticals*. He also conceives of the universe as involving a temporal series of world-wide episodes in which these substances participate. The hypotheticals provide explanations of each such episode relative to another. But Leibnoza, by accepting the principle of sufficient reason,

² Hegel did well to point out that the central concept of traditional rationalism was that of syllogisms *in re*. He also saw that the argument *in re* which, according to a thorough going rationalism, has as its conclusion the reality of *this* natural order rather than another—let alone the reality of *any* natural order—cannot itself be syllogistic in form.

demands in addition that every truth be either analytic or a necessary consequence of analytic truths.

Leibnoza, as a good Christian, believes that the world of interacting finite substances was created by God. And this means to him that God chooses to create *this* world rather than any of the other possible worlds which he might have created instead. It also means that this choice is in a relevant sense free. This freedom, however, must be compatible with the idea that there is a logically valid argument *in re* with a *logically necessary* premiss which proves the existence of this world. An impossible combination? Not for Leibnoza. He simply asks us to conceive of a set of possible Creators, each one freely choosing *sub possibilitatis* to create a different possible world. He then points out that one of these possible Creators must be the most perfect and necessarily exists. To use a Leibnizian (and Spinozistic) turn of phrase, the possible has a *misus* toward actuality in the sense that an *unimpeded* (or an insufficiently impeded) possibility is *ipso facto* actual. In short, what is logically necessary is not *the choosing* but *that the chooser of this choosing exist*. It is indeed logically necessary that the choosing exist, but no *existent* which is not defined in terms of the choosing logically implies the choosing. The existence of God necessitates the existence of the choosing but God is defined in terms of the choosing. In short, Leibnoza (like Leibniz) applies to God the latter's solution of the free will problem as it applies to Julius Caesar.³

Caesar's decision to cross the Rubicon was free in that (a) the objective of the decision is internally consistent in the way in which the objective of an (impossible) choosing to stand and sit simultaneously is not; (b) the choosing is not a logical consequence of any fact about Caesar which does not include the choosing; in particular, it is not a logical consequence of his prior state of mind. It is, however, a logical consequence of his nature, for his nature is simply a set of states of affairs which *includes* the state of affairs which is choosing to cross the Rubicon, and in no other sense can be said to constrain or necessitate the act.

As the existence of Caesar entails the existence of his free acts, so the existence of God entails the existence of His free acts. The difference is simply that Caesar exists by virtue of a choice made by God—whereas God exists as being the most perfect of a set of possible Creators. If we transfer

Leibniz' attempted reconciliation of freedom with the principle of sufficient reason to Leibnoza, we get the following account of how the existence of this world can be logically necessary and yet be one of many possible worlds. For according to Leibnoza, this world necessarily exists because the possible God who freely chooses it *sub specie possibilitatis* necessarily exists. From this perspective we can see that the important difference between Leibniz and Spinoza is not that Spinoza thinks that Caesar's crossing of the Rubicon is a necessary consequence of possible being whereas Leibniz' does not; but rather that Leibniz thinks that the relation of possible being to the crossing of the Rubicon is of the form:

The possible God who freely chooses to create the possible substance which freely chooses to cross the Rubicon necessarily exists.

III

A brief excursus on the classical problem of relations will set the stage for the next step in the argument. Suppose one thinks that the truth of

This leaf is green

requires that there be an item inhering in this leaf which is its greenness in the metaphysical sense of a dependent particular numerically different from all other greennesses, even of exactly the same shade, which inhere in other substances. Then relational predication immediately generates a puzzle. Consider

S_1 is R to S_2

If we treat this proposition as a special case of

S_1 is P

thus,

S_1 is R -to- S_2

and attempt to introduce a dependent particular which corresponds to this predicate as a particular greenness corresponds to "This leaf is green," we are faced by a dilemma

- (1) Is the dependent particular an R -to- S_2 ? This would seem to require that S_2 inhere in S_1 as being a part of the R -to- S_2 which inheres in S_1 .
- (2) Is the dependent particular an R rather than R -to- S_2 ? If so then it inheres in either (a) S_1 alone, or (b) both S_1 and S_2 or (c) neither S_1 nor S_2 . But not (a), for then the fact that S_1 is R to S_2 would be unaccounted for; furthermore, it would imply that S_1 could stand in the relation without

³ *Discourse on Metaphysics*, section XIII.

having a relatum. And not (b), for "an accident cannot have its feet in two subjects." Even if S_1 and S_2 could share an R , S_2 might cease to exist (thus, be destroyed by God) and we would be back with the absurdity of the previous alternative. And not (c) for particulars other than substances are *dependent*, i.e., necessarily inhere in substance.

Leibniz found an interesting way out of this dilemma. In effect, he adopts a modified form of the first horn. He accepts the principle that if

$$S_1 \text{ is } R \text{ to } S_2$$

is true, then there must be an R -to- S_2 inherent in S_1 , and he accepts the consequence that S_2 must be *in* S_1 . But he reinterprets these commitments in the light of the cartesian (ultimately scholastic) distinction between "representative" (or "objective") and "formal" being. Thus, the R -to- S_2 inherent in S_1 is interpreted as a representing of S_2 inherent in S_1 , and Leibniz, therefore, interprets the sense in which S_2 is a "part" of the R -to- S_2 inherent in S_1 as a matter of its being that which has objective or representative being in the representing which is the R -to- S_2 . According to this analysis, the truth of statements of the form

$$S_1 \text{ is } R_1 \text{ to } S_2$$

where R_1 is *prima facie* a real relation, rests on facts of the form

$$S_1 \text{ represents (in specific manner } M_1) S_2$$

where, needless to say, the manner of representation M_1 which *corresponds* to R_1 and makes this relational fact a phenomenon *bene fundatum*, is not what common sense has in mind when it uses the term " R_1 ."⁴ If it is objected that on the above account

$$S_1 \text{ is } R \text{ to } S_2$$

could be true even though S_2 did not exist, since non-existent substances can be represented, Leibniz would welcome this objection, but turn its edge by agreeing that the truth of the relational statement requires the actual existence of both S_1 and S_2 , and hence that the mere fact that S_1 represents a substance in the appropriate manner does not make the corresponding relational statement true. The substance represented must have formal as well as objective being, in order for this

to be the case. After all, his problem was to resolve the classical puzzle about relations, and this he has done, to his own satisfaction, by giving phenomenal relations between substances a metaphysical underpinning in which they have as their real counterparts acts of representing and mobilizing the distinction between the two modes of being which representables may have. Roughly, a true representation is one whose subject matter is a representable which, in addition to having "objective" being in the representation has "formal" being in the world.

IV

If we apply these considerations to causality, we can understand why Leibniz believes himself forced to interpret the fact that

S_2 is acted on by S_1 (e.g., S_1 by being in state ϕ *causes* S_2 to become ψ)

as involving, among other things, facts of two radically different types:

- (1) S_2 representing the fact that S_1 is in state ϕ
- (2) S_2 being caused by representing this fact to become ψ .

The first type of fact is a matter of the ideal relation of *truth* between a judgment and the actual state of affairs which makes it true. That in most if not all cases of the action of one substance on another the judgment is "confused," is a complication which can be overlooked for our present purposes. What is important for our purposes is that the ideal relation between a judgment and the fact which makes it true and, in general, between "ideas" and their "ideata" was conceived to be the logical product of two relations, one between the idea and a content, the *being* of which was its *being represented*, and the other between this content and the "external" object or state of affairs.

Notice, in the light of the preceding brief exposition of Leibniz' general theory of relations, that although

S_1 by being in state ϕ caused S_2 to become ψ

implies that both S_1 and S_2 exist, so that it would be nonsense to say " S_1 caused S_2 to become ψ but that there is no such thing as S_2 ," Leibniz can argue

⁴ Thus the statement, in the phenomenal framework of material things in space,

S_1 is linearly between S_2 and S_3

might have as its real counterpart something like

S_1 represents S_2 and S_3 more directly than S_2 and S_3 represent each other

where a representing of S_i is indirect if it is a representing of a representing of S_i .

that the existence of S_1 and its being in state ϕ is *causally* irrelevant to S_2 's being ψ and is relevant only to the *truth* of the representation which is the *vera causa* of S_2 's being ψ . Thus S_2 is, as far as the *relation* grounding each of its episodes in other episodes is concerned, as self-contained as Leibniz's world of interacting things. Both, however, are contingent and require a self-affirming premiss as their sufficient reason. It is, at least in part, the fact that truth is a perfection, which rules out the possibility that the universe might consist of S_2 and God, S_2 's representation of S_1 being a representation of something that doesn't exist.

V

We can sum up our results to date by saying that one line of thought which underlies Leibniz' thesis that the nature of an individual substance entails episodic as well as hypothetical facts about it, stems from his commitment to the principle of sufficient reason. Yet the argument is incomplete, for granted that episodes have a sufficient reason, and granted that this sufficient reason does not involve the actual being of other substances, it could still be argued that although hypothetical facts and episodic facts alike are grounded (not, of course, independently) in the First Cause, we are not thereby forced to count the episodic facts as elements in the nature of the substance. Why not continue, with Broad, to identify its nature with the hypotheticals, while granting that both episodes and hypotheticals are grounded in Necessary Being?

Part of the answer lies in the fact that we have been guilty of an anachronism in attributing to Leibniz the contemporary distinction between causal properties as general hypotheticals and occurrent states as the categoricals which co-operate with the former in generating further categoricals in an ontological two-step. The truth of the matter is that Leibniz, like most of his predecessors and many of his successors, interprets causal properties on the model of desires, plans, personal commitments. Thus, whereas *we* might be inclined to interpret the statement "Jones has a strong desire to go to New York" in terms of conditional facts about Jones, Leibniz thinks of a strong desire as a continuing series of episodes which tends to develop into a going to New York and will to develop if not impeded. Thus, to be more precise, he tends to think of the fact that S_2 would become ψ if S_1 were to become ϕ as a matter

of S_2 having the plan of becoming ψ if S_1 were to become ϕ . For becoming ψ is (*realiter*) doing something. And, having, *the plan* of doing A if B is (though the *plan* is hypothetical in character) itself a categorical fact about S_2 .

Actually, then, there is a sense in which for Leibniz all the fundamental facts about a substance are episodic facts. And consequently the notion of the nature of a substance as the law-like hypothetical which would provide an explanation of each episode relative to another episode, is ultimately replaced by the notion of the nature of a thing as that which *logically* explains each single episode. And, of course, the only way it can do this is by duplicating in some way the set of episodes which it is to account for.

Suppose we were to press Leibniz with the question: What is the mode of existence of the nature of a substance, and how is it related to the substance? I think that he would answer somewhat as follows. The nature of a substance is to be construed as its life-plan, and as such it has *esse intentionale* as the content of an abiding aspiration which is the core being of the substance. This connecting of *truths* about what a substance will do with an abiding plan of life raises in an acute form the problem of the relation of time to truth.

VI

That problems pertaining to truth play a central role in Leibniz' metaphysics is a familiar fact. Thus he supported the ideas that the nature of a thing includes once for all everything which can happen to it, and that the individual concept of a thing includes once for all everything that can happen to it—which so startled Arnauld—by considerations pertaining to truth. Like the argument from explanation, the argument from truth purports to show that there are entities, the sort of thing that would usually be called facts, themselves without dates, (though they are *about* dates), which account for the truth of true ideas about individual substances.

Thus, suppose the following statements, made today, are true

- (1) S_1 was ϕ_1 in 1957
- (2) S_1 is ϕ_2 now
- (3) S_1 will be ϕ_3 in 1959.

To make these statements is to say that one and the same individual subject was ϕ_1 is ϕ_2 and will be ϕ_3 . On the assumption of a correspondence theory of

truth, each of these statements corresponds to a fact; thus

- (1) corresponds to the fact that S_1 was ϕ_1 in 1957
- (2) corresponds to the fact that S_1 is now ϕ_2
- (3) corresponds to the fact that S_1 will be ϕ_3 in 1959.

and indeed (always on the assumption of the truth of the original statements) it is a fact that S_1 was ϕ_1 in 1957; it is a fact that S_1 is ϕ_2 now; it is a fact that S_1 will be ϕ_3 in 1959.

We say of an episode that it took *place*, *is taking place*, or *will take place*. But if something is a fact, it is a fact—not *was* a fact nor *will be* a fact. This is not quite true, for there are contexts in which “was a fact” and even “will be a fact” do make sense. But these are derivative uses in which, roughly, we are viewing someone else in the past or in the future as thinking of something as a fact. Where it is *we* who are thinking of something as a fact, the proper expression of this thought is always by the use of the *present* tense of “to be a fact.”

Now it is easy to move from the impropriety, in non-oblique contexts, of “it will be a fact . . .” and “it was a fact that . . .” to the idea that the “is” of “it is a fact that . . .” has to do with a timeless mode of being. (After all, one does not say “2 plus 2 will be 4” or “2 plus 2 was 4.” Are not facts like numbers?) In effect, Leibniz makes this move. To make it is to suppose that there is a timeless set of entities, i.e., facts, which are about what happens to a substance at different times, and such that it is by virtue of corresponding to these entities that our statements and judgments about the substance are true.

Before we take a closer look at this ontology of truth, let us notice that even if it were illuminating to say that every true statement is true because it corresponds to a timeless fact, this by itself would give no aid or comfort to the idea of the nature of a substance as a set of facts which make statements about its history true. For unless one has already denied the reality of relations, there would be many facts (i.e., relational facts) which have more than one substance as constituents. In the absence of Leibniz’ theory of relations, therefore, the ontological version of the correspondence theory of truth would support at most the idea that the only thing which has a nature, strictly speaking, is the universe as a whole. But, of course, Leibniz does have his theory of relations, and so his theory of truth does dovetail with the argument from explanation. It is important, however, to note that

even in combination with his theory of relations the argument from truth provides in and of itself no reason for calling the timeless set of facts about a substance its “nature.”

Returning now to the idea of the nature of a substance as the abiding life plan of that substance, we note that to say that the statement

S_1 will be ϕ_3 in 1959

is true because S_1 *now* plans to be ϕ_3 in 1959 is to lose the *prima facie* advantages of the simple ontological theory and embark on an uncharted course. On the other hand, the notion of a timeless aspiration would seem to be nonsense. But even if Leibniz was guilty of the category mistake of conflating the notions of timeless facts and life plans, the former notion undoubtedly guided his thinking. It is therefore appropriate to make the point, scarcely a surprising one, that this notion won’t do at all. To see that this is so, one simply needs to see that the “is” of

It is a fact that S_1 will be ϕ_3 in 1959

is exactly what it seems to be, namely the present tense of the verb “to be,” and that the “will be” of the that-clause is exactly what it seems to be, namely “will be” in relation to the present tense of the main verb.

But surely, it may be said, this “is” can be in the present tense only if it *could be* a “was”—which you have denied. But I have not denied it. I have pointed out that “was” is appropriate, but only where we are indirectly expressing someone else’s point of view. And this is the heart of the matter. For

It is a fact that S_1 will be ϕ_3 in 1959

is, in a very important sense a counterpart of

“ S_1 will be ϕ_3 in 1959” is a true statement

(not, of course, as Strawson has pointed out,

“ S_1 will be ϕ_3 in 1959” is a true sentence.)

To refer to the statement “ S_1 will be ϕ_3 in 1959” is always to refer to the relevant sentence as used at a certain time. And when I say

“ S_1 will be ϕ_3 in 1959” is a true statement
the time in question is *now*. If I say

“ S_1 will be ϕ_3 in 1959” *was* a true statement

the reference is to the use of the sentence at a time before now; and if I say

“ S_1 will be ϕ_3 in 1959” *will be* a true statement
the reference is to the use of this sentence at a time

later than now. We can now see why, if we limit ourselves to fact statements which express our own point of view *hic et nunc*, there is no place for

It was a fact that S_1 will be ϕ_3 in 1959

The point can be made by supposing someone to ask: Why doesn't

It was a fact that S_1 will be ϕ_3 in 1959
correspond to

" S_1 will be ϕ_3 in 1959" was a true statement.
as

It is a fact that S_1 will be ϕ_3 in 1959
corresponds to

" S_1 will be ϕ_3 in 1959" is a true statement

The answer is, of course, that to use the expression *"that S_1 will be ϕ_3 in 1959"* is to imply a reference to the *present* use of the sentence " *S_1 will be ϕ_3 in 1959,*" whereas,

" S_1 will be ϕ_3 in 1959" was a true statement explicitly refers to a past use of this sentence. Thus, if we limit ourselves to fact statements which express our own point of view—that is, if we limit ourselves to fact statements which do not occur in oblique contexts—all fact statements will be counterparts of

X is a true statement

and will imply a reference to a *contemporary* use of the sentence represented by " *X* ".

There is an obvious comeback to this argument. It runs as follows: you have been considering the sentence " *S will be ϕ_3 in 1959*" and it must be granted that this sentence is used to make different statements at different times. And it is reasonable to argue that given a that-clause constructed from this sentence, i.e., a *tensed* that-clause, the "is" of "it is a fact that that" must be in the *present* tense. But this reasoning no longer holds if we turn our attention to the sentences " *S is ϕ_3 in 1959,*" " *S_1 is ϕ_3 in 1958,*" and " *S_1 is ϕ_1 in 1957*" where these sentences have been detensed. These sentences, the objection continues, make the same statement whenever they are used, and consequently the statements

It is a fact that S_1 is ϕ_3 in 1959
It is a fact that S_1 is ϕ_3 in 1958
It is a fact that S_1 is ϕ_1 in 1957

no longer need to be construed as being in the present tense, save in that "timeless" use of the present found in "*2 plus 2 is equal to 4.*" The

answer to this objection consists in showing that the *contrived* sentence " *S_1 is ϕ_3 in 1959*" makes sense only as introduced in terms of tensed sentence-forms and amounts to the disjunction

S_1 was ϕ_3 in 1959 or S_1 is ϕ_3 in 1959 or S_1 will be ϕ_3 in 1959.

Now for the purpose of my present argument, it will be useful to lay down an oversimplified thesis to the effect that statements of the form

It is a fact that p

are simply another way of saying that

" p " is a true statement in our language

and, to get down to fundamentals, that statements of the form

that- p is a proposition or state of affairs

are simply another way of saying

" p " is a statement in our language.

Now if these theses be granted, it follows that to say that

The statement " S_1 will be ϕ_3 in 1959" is true
because

It is a fact that S_1 will be ϕ_3 in 1959
is like saying

we're here because we are here

The "because" is out of place because nothing is explained. There is indeed a proper because-statement in the neighborhood, but it must be formulated, rather, as follows

The statement " S_1 will be ϕ_3 in 1959" is true
because S_1 will be ϕ_3 in 1959.

If the truth of statements about substances requires no ontology of *facts*, the argument from truth collapses. Furthermore, the implication is unavoidable that insofar as the concept of the nature of a thing is a legitimate one, it can be formulated in such a way that it requires no use of "fact" which is incompatible with the schema

It is a fact that- p if and only if " p " is a true statement of our language.

This, I believe, can be done. Of more immediate concern, however, is what might be called the promissory note character of the idea that things have natures. This promissory note character is a pervasive feature of the statements we are in a position to make about the world. But before I expand on this theme, there is one more strand to be disentangled from the thinking which finds expression in the thesis that the individual con-

cept of a substance includes once for all everything which can ever happen to it. This I shall call the argument from proper names.

VII

Before exploring the cluster of reasonings which I propose to sum up by the phrase "the argument from proper names" I shall set the stage by some informal remarks which will show how the land lies with respect to my own views on the matter. As I see it, proper names of things and persons play an indispensable role in discourse. Although essentially related on the one hand to definite descriptions, and on the other to demonstratives, they are reducible to neither. The fact that any *single* name can be dropped and replaced by a descriptive phrase, at least in a specific context, should not deceive us into thinking that all names can be dropped from the language in favor of definite descriptions.

Again, the use of demonstratives presupposes that the speaker locates himself and his hearers in a common world of objects in space and time; and the framework in terms of which this locating is done is constituted by the use of names and definite descriptions of abiding things. But if the use of demonstratives presupposes the use of names and descriptions, the use of names and descriptions in their turn presupposes the use of demonstratives. The meaning of names and descriptions alike requires that I be able to recognize a named or described object as *this* object. No one of these modes of references is, so to speak, the foundation of reference, the Atlas which supports all the rest.

That proper names presuppose definite descriptions is scarcely controversial. To use a name "*N*" is to purport to be ready to make at least one statement of the form

N is the *f*-thing

Thus names presuppose the statement form "Something is *f*." But in their turn statements of the form "Something is *f*" presuppose that we have some way of referring to objects other than by making general statements. The idea that the Atlas of reference is bound variables is as mistaken as the once popular idea that it is demonstratives. The statement

Something is red: $(\exists x)(x \text{ is red})$

is a functioning part of language only because there are names and demonstratives to function in determinate singular reference.

It is not, however, my purpose to *argue* that descriptions and demonstratives presuppose names (and vice versa). I shall simply assume that this is so. For my concern is with the question, "Granted that names are an irreducible mode of reference, what are the implications of the idea that every individual thing is *nameable*?" For if anything is a central fact in Leibniz' metaphysics, it is that he clearly assumes that every substance is nameable, and I believe that the recognition of this fact throws a flood of light on his system.

Let us call the idea that every substance is nameable the "Principle of Nameability." The first thing to note about this principle is that it stands in a certain interesting relation to the principle of the identity of indiscernibles (or the dissimilarity of the diverse). To begin with, if names were shorthand for definite descriptions (which they are not), to stipulate that every individual is nameable would be to stipulate that

$$(x)(\exists y)(\exists f)[x = (y)f]$$

from which it follows that

$$(x)(y)[x \neq y \rightarrow (\exists f)(fx \& \sim fy)]$$

The important thing to see is that even if one doesn't *equate* names with definite descriptions, the same conclusion follows from the principle of nameability if one makes the related claim that every name has a *sense* which consists of one (or more) definite descriptions. For one can hold that *being the f-thing* is a *criterion* for being correctly called *N* without holding that "*N*" is *shorthand* for "the *f*-thing."

It is sometimes thought that the reason why it is incorrect to characterize names as shorthand for definite descriptions lies in the precariousness of our beliefs about the world which makes it advisable to avoid pinning a name down to only one definite description. In short, it is thought that there are *practical* reasons independent of vagueness, open-texture and the like for refusing to equate names with descriptions—for one can grant that the use of names rests on a fallible inductive footing which warrants a looseness in the connection of names with descriptions, without construing this connection on the model of logical shorthand.

In any event, I am going to stipulate for the purposes of my argument, that names have definite descriptions as their senses, in that definite descriptions serve as the criteria for the application of names. Statements about named objects, then, presuppose (in Strawson's sense) the truth of

statements affirming the existence of a unique descriptum. Because, thus construed, names presuppose states of affairs, one can appreciate why those who seek an ultimate mode of reference which involves no commitment to something's being the case either deny (with Wittgenstein in the *Tractatus*) that names have a sense—or equate the sense with the referent—or, with Quine, deny that naming is an ultimate mode of reference and seek to reduce it to the use of bound variables. But this notion of a presuppositionless mode of reference is but another manifestation of the idea that empirical knowledge has a foundation, other facets of which I have explored in "Empiricism and the Philosophy of Mind."

Now I am implying that Frege's conception of proper names—though not his theory of definite descriptions—is not only sound, but closer to the tradition than certain modern alternatives. In particular, I am implying that it is close harmony with Leibniz' treatment of names, though the latter nowhere develops an explicit theory of names along these lines. Assuming this to be so, the first point I wish to make is that Leibniz is clearly committed to the idea, indeed takes for granted, that every individual substance is *nameable*, and that this acceptance of the Principle of Nameability carries with it the principle of the Identity of Indiscernibles. For according to the above theory of names, the name of each properly (and not merely putatively) named substance will have as its sense a criterion which distinguishes its nominatum from all other substances. Let us call this *sense* the "individual concept" which the name "stands for."

Now this is exactly the sort of thing Leibniz means by the phrase "individual concept." But whereas *we* would think that the individual concept for which a name stands need specify only a few facts about the nominatum, for we think that a few facts suffice to single it out from other things, Leibniz interprets the individual concept associated with the name as specifying everything that the nominatum does or suffers throughout its entire career.

Why should Leibniz think that the sense of a proper name must include a complete description of the nominatum? The answer to this question is surprisingly simple once one realizes that Leibniz is concerned not with *our* names for substances—indeed as we have already pointed out he thinks that the so-called names we use are not really names at all but a peculiar kind of general term—

but with *God's* names for things. Thus the principle of nameability is the principle that every individual substance is nameable by God. If, now, we bear in mind the argument that the sense of a name must serve to distinguish its nominatum from all other substances, we see right away what is going on. For Leibniz simply takes it for granted that it makes sense to speak of naming possible substances! And it is by no means implausible that, though an incomplete description of an object may serve to distinguish it from all other *actual* things, only a complete description which pins the object down in all conceivable respects in accordance with the law of excluded middle can distinguish it from all other *possible* things. If it were to be granted that God has names for all possible substances, it would seem indeed that the individual concepts for which these names stand must be as Leibniz characterizes them.

Now even before we turn the cold light of analysis on the idea of names of possible substances, we can put our finger on an ambiguity in this conception. We have pointed out that the sense of a name serves to discriminate its nominatum from the other members of a set of mutually discernible substances. Now, granted that the set of all logically possible substances is a mutually discernible set, the question arises "Is the mutually discernible set *which is relevant to the naming of a possible substance* the set of all *logically* possible substances?" The point is an important one, for it is only if the former is the case—only, that is, if the names of possible substances are prior to any *more restrictive* principles of compossibility which build logically possible substances into possible worlds—that it would be true that their "individual concepts" must describe them exhaustively. If the set of mutual discernibles relevant to the naming of a possible substance coincides with the more restricted notion of a possible world having extra-logical coherence, the individual concept of a possible substance could discriminate it from other substances in its world without having to describe it completely; and its distinguishability from all possible substances in other possible worlds would follow from the distinguishability of the worlds. To take this line in applying the principle of nameability is to deny that there are any possible substances apart from coherent possible worlds.

Now it seems to be quite clear that Leibniz actually thinks of the individual concept of each possible substance as specifying its place in a system of mutually adjusting substances which develops in the orderly lawful way characteristic

of a possible world. In so doing, I shall argue, he has undercut his demand, insofar as it is based on the idea of a name, that the individual concept of a possible substance selects that substance in terms of a *complete* description. Instead, however, of supporting this criticism directly, I shall do so indirectly by turning my attention to the question, What sense is there to the notion of the name of a possible substance?

VIII

Are there such things as possible substances? That question is best approached by considering a familiar case for the negative. It rests on the idea that the primary use of the term "possible" is in such context as

It is possible that Tom will get well.

It rests, in short, on the idea that it is states of affairs rather than things which are said, in the first instance at least, to be possible or impossible. And, it is argued, the statement that such and such a *state of affairs* is possible presupposes the actual existence of the *things* with respect to which this possibility obtains. Thus, the possibility that Tom will get well presupposes that there actually is such a person as Tom.

It might be thought that this argument is self-refuting. How can one properly argue that there are no *possible* things on the ground that possible states of affairs concern *actual* things? Surely to speak of "actual things" is to imply that it makes sense to speak of "non-actual" or "merely possible" things. But of course, the argument didn't say that in order for there to be the possibility that Tom will get well, Tom must be an actual thing, that only that there must actually be such a person as Tom.

Having made plausible the idea that the primary use of "possible" is in connection with *states of affairs*, the case for the negative now grants that a derivative use of "possible" might be introduced in which one would speak of possible *things* in accordance with the following schema

There is a possible man in the corner = it is possible that there is a man in the corner.

In this stipulated sense there would be possible things—for to deny it is to deny that sentences such as "It is possible that there is a man in the corner" can ever be used to make a true statement.

Now I take it to be common ground that such sentences as "It is possible that Tom will get well"

and "It is possible that there is a man in the corner" are often used to make true statements. I also take it to be common ground that these statements are the blood brothers of *probability* statements and as such are statements in the *present* tense which imply a reference to evidence now "at hand" or "available." Just as the probability statement

Tom will probably get well

has roughly the force of

There is a balance of evidence at hand in favor of the statement

Tom will get well

so the possibility statement

It is possible that Tom will get well

has the force of

There is no conclusive evidence at hand against the statement

Tom will get well.

and just as when more evidence becomes available, it may become proper to say

It is still very likely that Tom will get well

so, if the evidence is unfavorable, it may become proper to say

It is no longer possible that Tom will get well

or

The possibility that Tom will get well no longer exists.

IX

Now if we have reason to believe that there is a man in the corner, we can properly say "let's call him Jack." In this case "Jack" has as its sense "the man in the corner." As long as we have reason to think that there was a man in the corner at that time, we have reason to think that "Jack" is our name for a man. On the other hand, as soon as we have reason to think that no man was actually there, we have reason to think that in the use we gave to it, the word "Jack" does not name anything. For a word in a certain use, say "Jack," to be a name for something there must be such a thing as is specified by the sense we have given to the word—there must be such a thing as Jack. This is the insight contained in the slogan: A name isn't a name unless it names something. If, therefore, we discover that there was no man in the corner, we are no longer entitled to regard "Jack" in that use as a name. To be sure, we can *now* use it as *short for*

"the man who was in the corner." But if we do so, we are no longer using this descriptive phrase as the criterion of a *name*, and we can no longer say that this phrase gives the sense of a name. In short, we cannot say that "Jack" as we are *now* using it is a name.

It would be obviously absurd to say

It is possible that there is a man in the corner:
Let's call him Jack.

If, however, we have made the move from

It is possible that there is a man in the corner
to

There is a possible man in the corner
we may not see the absurdity. For if it makes sense to say

There is a tall man in the corner, let's call him Jack

why shouldn't it make sense to say

There is a possible man in the corner, let's call him Jack.

The mistake is obvious, for the latter simply repeats the absurdity of

It is possible that there is a (unique) man in the corner; let's call him Jack.

In short, it is only if one construes "naming" as introducing an abbreviation for a definite description that one will regard it as proper to speak of naming in this connection. For one would then construe

It is possible that there is a man in the corner;
let's call him Jack

as a paradoxical way of saying

It is possible that there is a unique man in the corner; let's use "Jack" as short for "the man in the corner."

Notice, however, that even this could not be construed as naming a "possible man" whose status as possible was independent of epistemic vicissitudes. For, if "Jack is a possible man" has the sense of "It is possible that there is a (unique) man in the corner," as soon as new available evidence about the status of the corner at that time requires us to say "It is not possible that there was a man in the corner," we would have to say "Jack is an impossible man."

⁵ If one allows that "there is a man in the corner" entails "it is possible that there is a man in the corner," where "possible" is used in the epistemic sense—which does some violence to ordinary usage—then one should say "... the fewer *merely* possible objects and states of affairs it will admit."

X

This last point reminds us that the case for the negative has been built on what might be called the "epistemic" sense of the terms "possible" and "impossible." This sense is not, of course, independent of the "nomological" senses of these terms, but must not be confused with them. In the epistemic sense, statements of possibility are relative to evidence available to the speaker. They pertain to the world not as it is "in itself" but to the world as known by someone in some circumstances at some time. And, as a first approximation, we can say that the more evidence that is available concerning a spatio-temporal region, the fewer possible objects and states of affairs it will admit.⁵

If we mobilize the Peircean idea of an inductive community, a community consisting of ourselves and those who join us, and suppose that our remote descendants in this chain have evidence and principles which enable them to decide with respect to every earthly spatio-temporal region whether or not it contained a man, we could imagine them to say

At such and such places and times there were men; at such and such other places and times there were no men

and, by way of epistemic commentary on the latter statement,

It is not possible that there were men at the latter places and times

or, by an extended usage,

There were no possible men at those places and times.

The truth of these future statements is no more incompatible with the truth of *my*

It is possible that there is a man in the corner

or

There is a possible man in the corner

(given that there is no man in the corner) than the truth of the croupier's statement

It is impossible that the dice showed seven is incompatible with the truth of my prior statement

It is probable that the dice showed seven
(given that I know my dice to be loaded).

It is worth pausing to note that these philosophers who argue that determinism implies that the possible coincides with the actual are guilty of two confusions:

- (1) they are telescoping epistemic and nomological possibility into one concept;
- (2) they are fallaciously supposing that because it would be true for a demon who knows a cross-section and the laws of a LaPlacian universe to say (with respect to any time t) "it is not possible that the state of object O at time t should have been other than S ," my statement "it is possible that O is not S at t " must be false.

The fact that determinism implies that an ideal knower could make no true statement of the form "both p and not- p are possible," where the epistemic sense of "possible" is involved, does not imply that it cannot be true for imperfect knowers to say "both p and not- p are possible." If, speaking as convinced determinists, we say that "when you come right down to it only what *actually* happens is *really* possible" this simply expresses our sense of community with those ideal members of the republic of investigators, the concept of which is the regulative ideal of the life of reason.

XI

At this stage, we may imagine the opponent of possible-but-not-actual things to grant that the epistemic sense of "possible" permits the truth of statements of the form "it is possible that- p " where not- p is, in fact, the case; and also that correctly construed, sense can be made of such statements as "there is a possible man in the corner." But, he adds, it is only if indeterminism is true that these possible-but-not-actual things are anything more than expressions of human ignorance. And even if there are possible-but-not-actual things (in the epistemic sense) which are *not* relative to ignorance, they presuppose the actual existence of the known evidential context (objects and statement of affairs) which fails to pin them down. Mere possibilities in the epistemic sense cannot, in the nature of the case, be prior to actual existence in the sense required by Leibniz' system.

All this, however, would be readily granted by Leibniz. For he is committed to the view that *in the sense of the previous discussion* there are no possible

but-not-actual substances save in relation to human ignorance. For, according to Leibniz, God creates one of the possible worlds (in a sense of possible to be explored) and each possible world being a maximum set of *compossibles* it follows that there are no states of affairs compossible with but not included in a given possible world to be the careers of substances which are possible-but-not-actual relatively to that possible world. Consequently, whichever possible world is the actual one, it could only be in relation to incomplete evidence that a knower in that world would be entitled to say "it is possible that there is a man in the corner" when in point of fact there is not.

Now given a set of qualities and relations and supposing all simple qualities and relations to be compatible, i.e., supposing no extra-logical limitations on compossibility, it is not possible to define more than one maximum set of compossible objects.⁶ In short, it is not possible to define more than one maximum world involving these qualities and relations. (And what a queer world that would be!) This means that in order for there to be more than one maximum world of compossible substances, an additional restriction must be introduced. And this additional restriction pertains to the coherence of possible worlds. The way in which Leibniz introduces this additional restriction is instructive. He introduces a reference to the decrees of the Creator into the defining traits of a possible world. After all, since the proximate possibility of a possible world is the possibility of the act of creation by which it would come into existence, the logical consistency of the world is but the possibility of this proximate possibility. Leibniz tells us that the possibility of *choosing* to perform a certain act presupposes that the description of the act is not self-inconsistent. To illustrate: the impossibility of *choosing* to stand and sit simultaneously stems from the impossibility of standing and sitting simultaneously. On the other hand, simply because a state of affairs is self-consistent, it does not follow that it is possible that I choose it. If, therefore, we assume that extra-logical modes of coherence are perfections, there would presumably be more than one maximum system of objects which a possible creator might choose to create, only one of which, however, could be chosen by the best possible creator, who necessarily exists. For, as we have seen, the possible, the actual, and the necessary coincide *in the last analysis* for Leibniz as for

⁶ Objects must be carefully distinguished from states of affairs (possible facts). Among the latter are negative states of affairs, and there would be many systems of states of affairs (state descriptions) with respect to the objects in question.

Spinoza. In this sense the actual world is the only one that is *really* possible.

XII

We have seen that Leibniz' notion of possible-but-not-actual things cannot be justified in terms of the inductive or epistemic use of "possible," and this for two reasons: (1) there is no such thing as naming possible objects in any sense of "possible" springing from this use; (2) a commitment to determinism involves a commitment to the idea that "the *really* possible coincides with the actual."

Is there any other line of thought which enables us to understand what was at the back of Leibniz' mind? The answer lies in the fact that it does make sense to speak of the *actual* world, and, by contrast, possible-but-not-actual worlds. Now it might be thought that to speak of "the actual world" is to refer to the actual course of *the world*, as contrasted with possible-but-not-actual courses of *the world*. If this were the case, the possibility involved would be analogous to the possibility that Tom will get well, and would be either the inductive possibility expressed by

It is (on the evidence at hand) possible that Tom will get well

or would be a misnomer for the *contingency* of the state of affairs—i.e., the fact that the idea that Tom will get well is neither analytic nor self-contradictory. But as we have seen, both the possibility that Tom will get well and the contingent character of this state of affairs presuppose the existence of Tom and of the world which includes him.

But what can the contrast between the actual world and possible-but-not-actual worlds amount to if it is neither of these? The general lines of the answer emerge if we remember that to think of the *actual* world is to think of what it would be *known* as by our ideal inductive decedents, that is, by ideal members of *our* scientific community. *This* world is essentially *our* world. *This* is what I am perceiving, and can talk about with *you*. What, then, does it mean to conceive another world? It means to conceive not only of a different set of substances than those which make up this world, but a set which does not include *us*.

Now if to think of an object is to use an expression which refers to it, we can't think of these other substances by using the referring expressions of the language we actually use. They refer *ex hypothesi* to

the objects of the actual world. It would seem, then, that to think of the objects of a non-actual world, we would have to use a language we don't have. Again, one cannot specify what object a referring expression refers to save by *referring* to it by means of another referring expression. And this referring expression would be one which belongs to the language we use and which talks about "actual things."

Fortunately there is a simple model at hand for talking coherently about non-existent things, and using names which don't really name anything. This model is fictional discourse. Thus, to "imagine" that there was such a person as Oliver Twist is to do what we would call purporting to describe an actual person, if the description didn't occur in a rubric ("once upon a time") which brackets discourse in a way which frees it from responsibility to inductive confirmation.

Thus to imagine a possible but not actual world is to place discourse which, if *seriously* intended, would purport to describe this *world*, in a rubric which marks it as fiction. And, of course, the "names" in this fictional discourse are only *fictional* names, and the sentences which constitute their *principia individuationis* are fictional sentences. The referring expressions in this fictional discourse do not translate into the language we seriously use. Language about a possible world is not a *different* language simply in the sense in which Italian is different from English—for "*Parigi*" does translate into "Paris" and we can say that "*Parigi*" refers to Paris.

To say what a language about a possible world is about we have to make *fictional* use of another language into which the *fictional* object language translates.

Furthermore, to *imagine* a possible world is to imagine that "I" belong to this world and am a member of an inductive community ferreting out its secrets. For the concept of *actuality* includes a reference to us and ultimately—if ungrammatically—to "I". Consequently, imagined actuality involves a reference to an imagined "I".

Is it not self-contradictory to speak of imagining that I am in a different world? Am I not, by definition, in *this* world? No, for to imagine that I am in a different world is simply to suppose that the states of affairs by virtue of which the general criteria associated with the term "I" are satisfied are other than they actually are; it is to suppose that these states of affairs are among those which belong to the fictional rubric.

But this is not the place to ferret out the truth in Kant's conception of the transcendental unity of appreciation as the fundamental form of experience. Our task is rather to illuminate Leibniz' conception of the actual world as one of an infinite number of possible worlds. What I am suggesting is that at the back of Leibniz' mind is the picture of God as making use, within the fictional rubric, of alternative languages, and by so doing conceiving of alternative sets of individual substances.

According to this picture, the model for *creation* is obviously the removing of the fictional rubric from one of these languages; the move, on God's part, from "Suppose that there were such and such things" to "There are such and such things," *via* "Let there be such and such things."

XIII

Now it is obvious that if the *esse* of possible worlds like the *esse* of Oliver Twist consists in fictional discourse about them, we have a new sense in which *actuality* is prior to possibility—roughly that in

which Dickens is prior to Oliver Twist. This accords with Leibniz' contention that the Divine Understanding is the locus of possible world. How, then, can he extend the notion of possibility to God Himself? For God is that Being who necessarily exists if He is possible. And God *qua* possible can scarcely have a being which is dependent on God's Understanding. I shall limit myself to two points:

- (1) When Leibniz tells us that the Divine Understanding is the locus of the possible, he seems, in the whole, to have the contingently possible in mind, and to be telling us that the *real* or *proximate* possibility of the contingently possible lies in the possibility of its intended Creation, and hence involves a reference to God's thought of it. This is compatible with an account of possible substances which make them prior to the actuality of God.
- (2) If my positive argument is correct, the actuality of God as of anything else would presuppose *our* existence as discoverers of God.

University of Pittsburgh

III. VAGUENESS, MEANING, AND ABSURDITY

HAIG KHATCHADOURIAN

I

THE logical positivists' or logical empiricists' celebrated—or should I say notorious—verifiability theory of meaning is defective or inadequate in a number of ways, for a number of fundamental reasons. This is true even with regard to its most defensible form or forms; e.g., in the form in which A. J. Ayer advocates it in his *Language, Truth, and Logic*. At the time I am writing, this is hardly news to many philosophers. One of the aims of this paper, however, is to point out and discuss one important failing or limitation of this principle which has tended to be ignored by philosophers. This particular limitation, apart from its logical or theoretical significance, is of very considerable practical importance because it arises, in philosophy, in relation to many traditional philosophical utterances, "doctrines," "theories," or "views" that purport to be empirical in nature. It is this latter aspect of the matter which gives our discussion much of its philosophical interest or importance, its philosophical point.

The failing or limitation of the verifiability theory of meaning I have in mind concerns a certain class of utterances, very common in traditional philosophy, which purport to express true propositions but which are, or may be, both non-confirmable and non-disconfirmable. Some of these utterances purport to express true empirical propositions; but they are in point of fact empirically neither confirmable nor defeasible. Yet these utterances are not devoid of all meaning in the

relevant ordinary sense of "meaning"; they do have a signification¹ (what I shall henceforth refer to as "meaning₁") as a whole. That is, these utterances are meaningful in the very sense of "meaning" in which the verifiability principle denies meaningfulness, regards as nonsensical, non-testable utterances that purport to express empirical propositions.² Those utterances, of the kind I have in mind, that purport to express analytic or necessary propositions, are likewise meaningful as a whole even though they cannot be shown to be either necessarily true or necessarily false by appeal to the customary meanings₁ of the words that compose them and the syntactical relations that obtain between them.

The utterances I am referring to are simply those which we ordinarily call vague statements; utterances which express vague propositions. (But see Sections II and V). For the purposes of this paper I shall appeal chiefly to the reader's knowledge of the ordinary meaning or uses of the word "vague" as applied to words, phrases, sentences, statements, or propositions. But I shall state, rather briefly, the main results of my analysis of the ordinary concept of vagueness as it arises in relation to words or phrases (and therefore also concepts), in my "Vagueness."³ I shall also point out the application of these results to vague statements and propositions, which are our chief concern here.

To begin with the negative thesis of that paper, it should be noted that, contrary to the widespread belief among philosophers, vagueness in the

¹ In the sense of "that which is intended to be or actually is expressed or indicated. a. Of language, a sentence, word, etc.: The signification, sense, import; a sense, interpretation." *The Oxford English Dictionary*.

² Unless the positivist makes it self-contradictory for an utterance of this sort to have meaning₁, by defining "meaning₁" in terms of testability. But that would defeat its own end. It would put the carriage before the horse; since a statement cannot be meaningless₁ as a consequence of being untestable: it can only be untestable as a consequence of being meaningless₁ (assuming that this is the only possible cause of untestability). For it can be shown—indeed, this would constitute the most fundamental objection, or ground for objection, to the verifiability principle—that if, or insofar as, the logical positivist defines cognitive meaningfulness (meaning₁) in terms of testability, he drastically modifies the ordinary meaning₁ of "meaning" or even replaces it by a new, philosophical meaning. On the other hand, if the logical positivist regards the verifiability principle merely as a criterion and not as a definition of meaning₁, he cannot reject out of hand the thesis and arguments concerning vague statements or propositions presented in these pages. He cannot appeal to an unusual definition of meaning in terms of testability (i.e., his own) as a basis for rejecting our thesis. At the same time other, though less serious, objections may be urged against this more moderate interpretation of the verifiability principle.

³ *Philosophical Quarterly*, vol. 12 (1962), pp. 138–152.

ordinary meaning of this word is distinct from, and dissimilar in many important respects to, the "borderline" or "marginal indeterminacy" of many, if not all, ordinary and scientific expressions; or to what F. Waismann⁴ calls the open texture of ordinary and scientific concepts. To use Max Black's definition, an expression that involves marginal indeterminacy is one in whose case "borderline cases" or "doubtful objects" are easily found. . . ."⁵ C. S. Peirce,⁶ Bertrand Russell,⁷ Max Black,⁸ Carl Hempel⁹, and Irving Copilowish (Copi)¹⁰ are only a few of the many recent or contemporary philosophers who confuse or mistakenly identify the two notions. Relatively few ordinary expressions are vague; the vast majority of these expressions are perfectly precise in an ordinary sense of "precise." On the other hand most or all ordinary expressions do involve some degree of "marginal indeterminacy," or the concepts they express are open-textured. Thus Hempel is wrong in holding that "no term of any interpreted language is definitely free from vagueness. . . ."¹¹ Similarly Bertrand Russell is wrong in holding that "all language is more or less vague."¹² It is not even true, as Black holds in his paper, that "all 'material' terms—terms whose application requires the recognition of the presence of sensible qualities—are vague; though many or all of them involve a 'fringe' or borderline cases in their customary application. The words 'chair,' 'stool,' 'pencil,' 'green,' and 'hard' are a few examples of this." (In his later "Additional Notes and References" in *Language and Philosophy*, p. 170, Black appears to agree with Hempel that both descriptive and logical terms in ordinary language are subject to vagueness.)

On the positive side, it should be noted that the applications of vague expressions in general, unlike

many or most of the (correct) applications of expressions that involve marginal indeterminacy, are not governed by any relatively-fixed and well-defined linguistic rules of a *specific* (or sufficiently specific) nature. There are, and there can be, therefore, no paradigm instances of the application of vague expressions. Or these expressions cannot be properly said to express some one or more well-defined (and sometimes not even fixed) concepts or ideas. But some of them—e.g., those that are nouns or adjectives—do express some concept or concepts. This is true even with respect to extremely vague expressions; but the difference between an extremely vague concept and no concept at all is extremely slight.

Because the uses of a vague expression are not governed by well-defined conventions, its applications by different persons, or even by the same person in different contexts, tend to fluctuate to a lesser or greater extent. However, these different applications may, nonetheless, involve a "common core" of meaning, differing in extent with different vague expressions, in a given period of time. Some or all of the meanings that are given to it in its various applications may partly overlap. This is generally if not exclusively true, I think, with regard to slightly or moderately vague expressions. In the case of extremely vague expressions, what little content the concepts they express (if any) have may be so hazy or indefinite as to make it meaningless or virtually meaningless to speak of them as possessing or lacking a "common core" of meaning in their different applications.

It is true that there is, or may be, some degree of flexibility or fluidity in the ordinary applications of perfectly precise (non-vague) expressions, whether or not they are open-textured. But all such (permissible) fluctuations of usage (and sometimes, also

⁴ "Verifiability," *Logic and Language*, first series, edited by Antony Flew (New York, 1951), pp. 117–144. Although Waismann characterizes open texture somewhat differently from the way in which Black characterizes marginal indeterminacy, both seem to me to be referring to the same phenomenon. It is noteworthy, however, that Waismann distinguishes vagueness from open texture. The latter, he holds, ". . . is something like the possibility of vagueness." (p. 120).

⁵ "Vagueness: An Exercise in Logical Analysis," *Language and Philosophy* (Ithaca, New York, 1949), p. 33.

⁶ *Baldwin's Dictionary of Philosophy and Psychology*, vol. II (1902), p. 748. Peirce's definition of vagueness is quoted by Black, *op. cit.*, p. 30.

⁷ "Vagueness," *Australasian Journal of Philosophy*, vol. 1 (1923), p. 88.

⁸ *Op. cit.*, pp. 25–58. Also, "Additional Notes and References," pp. 249–250.

⁹ "Vagueness and Logic," *Philosophy of Science*, vol. 6 (1939), pp. 163–180.

¹⁰ "Border-Line Cases, Vagueness and Ambiguity," *Philosophy of Science*, vol. 6 (1939), pp. 181–195. In this paper Copilowish first mistakenly identifies vagueness with marginal indeterminacy; then, later on, commits the further error of suggesting that "the notion of vagueness can be reduced to the status of a special but important case of ambiguity. . . ." (p. 181) For the main differences between vagueness and ambiguity, see my "Ambiguity," to appear in *Theoria*. Max Black, who rightly distinguishes vagueness from ambiguity, vitiates this distinction by his identification of the former with marginal indeterminacy.

¹¹ *Op. cit.*, p. 170. Cf. also pp. 163, 177 and *passim*.

¹² *Op. cit.*, p. 88. Quoted from Black, *op. cit.*, p. 26, n. 5.

of use) are always minor or slight, relatively to the extent of the uniformity or fixity in the usage (and uses) of these expressions. They all occur within the framework of, or are limited by, the rules governing their customary usage. No such drastic limitation, in principle, is imposed on the applications of vague expressions; since, as we said before, there are no well-defined and relatively-fixed conventions of a specific nature (i.e., other than very general or relatively general conventions) governing their usage in general. In the case of an open-textured expression, the fluctuation often arises with respect to marginal cases; though some of its users may stretch or narrow down its uses in relation to non-borderline objects. Paradigm cases obtain with respect to these expressions, just as they obtain with respect to expressions that are not open-textured. Or the *range* of actual or possible fluctuations in the applications of open-textured expressions is very much less than in the case of vague, particularly very vague, expressions. Moreover, the meanings or senses of the former do not change or vary even when the same expression is applied to a particular *marginal* object by one person and withheld by another, whereas the meaning of a vague expression may fluctuate or change entirely in its different applications.

The application of the foregoing outline of the ordinary notion of vagueness, as it arises in relation to words, phrases, and concepts to a discussion of statements and propositions, is clear. A statement which contains one or more vague expressions will frequently be a vague statement. (For the reason why I say "frequently" rather than "always," see under Section V). Consequently the proposition expressed by such a (vague) statement will be a vague proposition. But the analysis of vague concepts does not carry us all the way. For not every vague statement is necessarily one in which a vague expression occurs; and therefore not every vague proposition is one expressed by such a statement. As we shall see a little later, the vagueness of some vague statements has a different source or cause. Similarly, *mutatis mutandis*, with some vague propositions.

The negative part of our thesis in "Vagueness," summarized earlier, is of equally great importance for our present discussion. For the confusion of vagueness with marginal indeterminacy entails, among other things, the confusion of vague statements with utterances about borderline cases of the application of some ordinary expression or expressions—which, unlike vague statements, have

only the appearance of being statements. Like genuinely vague statements, these utterances cannot be properly said to be either true or false; but only because they do not state or assert anything, do not make any propositions about the things to which they refer. For they are really covert recommendations, or expressions of an implicit "decision" to use a particular word or phrase—the word or phrase which expresses an open-textured concept—in a new way, cast in the form of indicative sentences. Thus the utterance, "Viruses are living organisms," which is not vague at all but is the expression of an implicit "decision" to extend the application of "living organism" to viruses, has the air of a genuine statement, affirming something true or false about viruses. As with vague statements, such utterances as "Viruses are living organisms" are incapable of confirmation or disconfirmation; but only because they do not express any propositions at all. Consequently they do not, in contrast to vague statements, constitute exceptions to the logical positivist's verifiability principle (assuming that this principle is sound in other respects) or in any way affect its validity or invalidity.

Wherever vague statements occur—in mathematical, in scientific, or in ordinary discourse, in philosophy—the Humean-type dichotomy between (*a priori*) analytic and (*a posteriori*) synthetic propositions, on which the verifiability principle itself partly rests, becomes inapplicable. Vague propositions are neither analytic nor synthetic, properly speaking, though they may purport to be either the one or the other. And in their case, as I stated before, the verifiability principle's own dichotomy, relating as it does to *bona fide* synthetic propositions: "Either testable and (hence) meaningful₁, or untestable and (hence) meaningless₁, nonsensical," cannot apply.

II

In Section I, I gave a brief analysis of some of the uses of "vague" as it applies to words, phrases, and concepts and, consequently, also to statements and propositions. I shall now consider a second major sort of use—perhaps a second sense—of this word in relation to questions, statements, propositions, thoughts, beliefs, opinions, or views as well as words, phrases, and concepts. In the present uses of "vague," the vagueness of a vague word or phrase is not due to its lacking relatively-fixed and well-defined uses of meaning₁. Rather, it is due to

its being too general or not sufficiently specific in meaning₁. I say "too general" or "not sufficiently specific," not "very general" or "not specific" because—and this is the crucial point—it either fails to pinpoint its referent or purported referent, or its signification is too inclusive or insufficiently exclusive. Consequently statements or questions in which it occurs will tend to be too general for the purpose or purposes for which they are intended. They will thus be vague in the sense that they will fail to make clear (or definite) what exactly or precisely they are, or are intended to be about. And this means either (a) that nothing precise or specific is referred to, that they fail to make clear what particular object or class of objects, state of affairs, etc., they are intended to refer to; or (b) that nothing sufficiently specific or definite is said about their intended referent; that they fail to make clear that which they intend to say about it; or both (a) and (b). Let me illustrate these things by a trivial but fairly clear example. The words "positive" and "control" (also "controlled") are perfectly precise or non-vague in their ordinary uses. But the phrase "positively controlled," at least in the statement "Headache is positively controlled by *X* (the name of a medicine)," is rather vague. This statement, because of the non-specificity of "positively controlled," does not make it clear whether *X* is supposed to stop headaches completely or only to some extent (and if the latter, to what extent). Similarly with "Dandruff is positively controlled by *K*," and like statements. In contrast to this, the word "internationalism" is vague in the sense that, as a word, it lacks a relatively fixed and well-defined meaning₁ or a number of such meanings₁. However, as a logical consequence of this it fails (as the phrase "positively controlled," *mutatis mutandis*, fails) to refer precisely to whatever attitude, philosophy, or state of affairs it may be intended to refer to in particular statements.

Now compare the foregoing with the following type of situation. A sentence which, in some contexts, would *not* be normally regarded as expressing a vague proposition at all in the present (and also in our first) sense of "vague," may nonetheless be said to constitute a vague *answer* to some particular question of a more specific character than it. Thus "Many of the inhabitants of Africa belong to the Hemetic race" is a vague answer to, say, the question, "What proportion of the inhabitants of Africa belong to the Hemetic race?" Similarly, "Africa is a large continent" is a vague answer to the question, "What is the area of Africa in square

miles?" though it is certainly not a vague statement in many other contexts. The former statement is vague as an answer to the particular question because it fails to locate precisely the *referent* which the question is inquiring about; while the latter statement is vague in the present context because it fails to convey the desired *information* about the area of Africa (the referent) in terms sufficiently specific for the purposes of the question asked. Some descriptions are vague in a given context because they are expressed in terms of determinables rather than determinate or relatively determinate qualities, properties, or relations. An example is: "It is a small, colored glass object" given in answer to the request, "Describe the object (e.g., a small crystal vase) before you." A statement or passage in which both the reference and the information conveyed are not sufficiently specific will be very vague. But a statement may be quite—sometimes, extremely—vague even if only one or the other is not sufficiently definite. Degrees of vagueness arise in relation to each of the two separately, as well as the two together.

We should note here that statements of the sort under consideration *are* either true or false, analytic or synthetic. In other words, what we said in Section I about vague statements does not apply to them: it applies to statements that are vague by virtue of containing one or more expressions that are vague in our *first* sense of "vague." It does perhaps also apply to statements that are vague by virtue of containing one or more expressions that are vague in our *second* sense of the word. But we shall not attempt to determine whether or not this is the case.

III

In the foregoing discussion I have frequently used the words "definite" and "indefinite." It is useful at this point to say a few things about the uses of these words in relation to vagueness.

There are a number of senses of "definite" and "indefinite" that are relevant here. The first is the sense in which "vague" and "indefinite," "precise" ("non-vague") and "definite," respectively, are synonymous. Or, rather, insofar as there are two different senses of "vague" and "not vague," the sense of "indefinite" (and "definite") correspondingly differs as the one or the other sort of vagueness is in question. Thus on the one hand we say that the phrase "positive neutralism" is a vague or indefinite phrase; and on the other hand we say

such things as: "What you've just said is very indefinite, very vague"; "He has a very vague or indefinite idea about the facts of the case"; or "He is indefinite or vague about the significance of recent events." We also speak of a plan, scheme, or course of action as being indefinite or vague; as in "Our wedding plans are still very indefinite, vague."

There are, in contrast to this, a number of uses or senses of "definite" and "indefinite" in which these words are used very differently from "vague" and "not vague" respectively. The following statements illustrate some of these different uses: (a) "His trip to Europe has been postponed to the indefinite future"; (b) "Give me a definite date," or "There is a definite time, every year, when the swallows fly South"; and (c) "She has definitely made up her mind." Corresponding to use (b) of "definite" there is a use of "precise" in which the latter is *not* synonymous with "not vague." (There are no uses of "precise" corresponding to uses (a) and (c) of "definite.") I might add here that there is a further sense of "precise" in which this word is not synonymous with "not vague"; I mean the sense in which it refers to the *closedness* of some expressions or concepts. In this sense mathematical or logical expressions and concepts are precise; while many or all ordinary and scientific expressions and concepts are imprecise (open-textured). However, I am not sure that this sense of our expressions is not a technical one, invented by mathematicians, logicians, or philosophers. In any case, "imprecise" in this sense is not synonymous with "vague."

There is a further sense in which we speak of *statements* and *propositions* as "definite" or "indefinite." In this sense we speak of such statements as "Some forms of cancer are caused by heavy smoking" and "Many people are unaware of the recent advances in nuclear physics"—in general, statements that involve the use of "some," "many," "most," "a few," "the majority of," and so on, in the appropriate places—as indefinite.¹³ The indefiniteness of these and similar statements is quite different from that of such (vague) statements as "Dread is pervaded by a peculiar kind of peace. . . . Although dread is always dread of, it is not dread of this or that."¹⁴ But these and similarly indefinite statements may be vague in some contexts, in our *second sense* of "vague," just like "Africa is a large continent," about which we spoke earlier. Indeed,

our other example there, i.e., "Many Africans belong to the Hemetic race" is, *qua* statement, indefinite in precisely the sense with which we are here concerned. Thus in those contexts in which this statement is vague, it will be indefinite in *both* senses of this word: it will be indefinite in the sense of "vague" as well as in the sense of "not referring to any specific part of the particular class spoken about," or "not referring to any specific member or members of the class spoken about." Analogously, "Some mystics experience union with Ultimate Reality" is both indefinite in the latter sense and vague (in our *first* sense). It is, therefore, also indefinite in the sense in which "indefinite" is synonymous with "vague."

IV

Let us now consider some of the relevant uses of "absurd" and "not absurd" in relation to our discussion of meaning₁ and vagueness. These words are not ordinarily applied to individual words, phrases, or concepts but are confined to such things as statements, ideas (in the sense of propositions), views, doctrines, and theories. Thus the phrase "square circle," or even "emotional table," is not ordinarily said to be absurd (or not absurd). It may be that we would normally say that (a) the combination, in the sense of the conjoining or combining of "square" and "circle," or "emotional" and "table," in the present way, and even (b) the conjunction or combination itself, "does not make sense." If so, it only means that the ordinary notion of "not making sense" is broader than the notion(s) of absurdity.

There are, I think, three senses or sorts of uses of "absurd" relating to verbal expressions; and a fourth (and perhaps also a fifth) sense of it relating to actions, occurrences, situations, or states of affairs. Thus (1) we commonly use the word in the sense in which a *statement* such as "Friday (the fifth day of the week) is in bed," or the *proposition* expressed by it, is ordinarily said to be absurd. (2) We also use the word in the sense in which the *supposal*, *belief*, *view*, or *theory* that our lives are governed by the stars or that people in the antipodes are upside down would, or may be nowadays regarded as absurd. I am also inclined to think that (3) we speak of *obvious* or *patent self-contradictions*, such as "John is six feet tall and not six feet tall" or

¹³ Cf. M. Black, *Critical Thinking* (New York, 1950), pp. 169–170. Concern on the part of logicians with this sort of indefiniteness goes back to Aristotle, *De Interpretatione*, I, 7.

¹⁴ Martin Heidegger, "What Is Metaphysics?" in *Existence and Being* (Chicago, 1949), p. 366.

"Statement *S* is both wholly true and wholly false" as absurd, if and when they are, or we think they are, seriously asserted by someone who understands the meaning₁ of the statements, is in his right mind, and so on. It may be, however, that we apply the word only to the speaker's or writer's serious *assertion* (asserting) of them or even his serious *consideration* (considering) of them as possibly true statements, not to the latter themselves or the propositions they express. (4) There is, finally, a sense of "absurd" which does not really concern us here since it arises in relation to actual or possible occurrences, actions, situations, attitudes, or states of affairs rather than verbal expressions. Thus we sometimes say: "He put himself in an utterly absurd situation (position) by his thoughtless actions"; "What an absurd (silly, foolish) thing to do!"; "Going to bed in a formal dress! This is the height of absurdity!"; also, "Unreasoning hatred is absurd," "Blind jealousy is absurd," and "War is utterly absurd." Gilbert Ryle is therefore wrong in holding that "Only expressions can be affirmed or denied to be absurd. . . . For what is absurd is unthinkable"¹⁵; though he is certainly right in holding that "Nature provides no absurdities. . . ."¹⁶

We shall now consider senses (1) and (2) in some detail.

Sense (1). We ordinarily say that the statement "Friday (the fifth day of the week) is in bed" is meaningless. And by this we mean that it "does not make sense," lacks a *coherent* meaning₁, is absurd (hereafter referred to as "absurd₁"); though its vocabulary is conventional and it is syntactically well-formed. But it is not a bit vague. However, absurd₁ statements, just as non-absurd₁ statements, may be vague in some degree or other. This does not contradict our earlier statement that a vague expression or statement is meaningful₁. Yet how can a vague statement be absurd₁ and so not make sense; or how can an absurd₁ statement be vague and so be meaningful₁? The answer appears to be the following:

The words "meaningful" and "meaningless" are used in two distinct senses or meanings in the two cases. The sentence "Friday is in bed" becomes absurd₁ if, in a given instance of its use, "Friday" is

taken to mean a (certain) day of the week. But even then it will have some meaning₁ *as a whole*. As a matter of fact it could not be absurd₁ if it did not have any meaning₁ as a whole. "Abradacabra is brillig" or "Prat pars is panatar peewi" are meaningless₁ or nonsensical as a whole (in these particular instances, the individual components, with the exception of "is," are also nonsensical). In contrast to this, "Friday is in bed" does have some meaning₁ as a whole even in its absurd form (note the last phrase). To state the matter in other and more precise logical terms, an absurd₁ statement is a statement that commits a type-mistake. The day of the week called Friday is not the type, sort, or kind of thing, in an ordinary sense of these words, that can possibly (logically speaking) be either in bed or not in bed. Or stated in semantic terms, the word "Friday" as used to name a day of the week rather than, say, a person or an animal, cannot be conventionally coupled with "in bed" or "not in bed" in their relevant ordinary meaning, or vice versa. A linguistic rule or convention, or a "type," is violated if the two, in the sense or use indicated, are coupled in the above manner. Following Fred Sommers,¹⁷ I shall speak of such an expression-pair as "*N* in sense value." And I shall speak of an expression-pair which can be coupled without absurdity in some sense or senses they conventionally possess—e.g., "Friday" as the name of a man or animal, and "in bed"—as "*U* in sense value."

I said that an absurd₁ statement is one which commits a type-mistake. But only some statements that commit a type-mistake are *obviously* absurd and so their absurdity does not need to be exhibited by analysis or argument (*a reductio ad absurdum* or indirect argument). As a matter of fact, a statement which involves a type-mistake may be mistaken for a true (hence perfectly sensible and self-consistent) statement. "Time is the duration of human selves" and "Time has no beginning" are perhaps examples of this. At the same time these statements, just like statements in whose case the existence of a type-mistake is obvious, *may be* self-contradictory. I say "may be" rather than "are" self-contradictory because some absurd₁ statements are, while others are not and cannot be, self-contradictory. An

¹⁵ "Categories," *Logic and Language*, edited by Antony Flew, Second Series (Oxford, 1955), p. 76.

There is some reason to suppose that in the statement "Life is irrational, meaningless, absurd" the word "absurd" is used in a sense distinct from, though in some ways analogous to the above, fourth sense of the word. If this is true, we here have a *fifth* sense of "absurd" and "non-absurd." For at least part of what those who speak of life as absurd mean by it is that life lacks an ultimate rational goal or purpose.

¹⁶ *Ibid.*

¹⁷ "The Ordinary Language Tree," *Mind*, vol. 68 (1959), pp. 160–185.

example of the former is "Time is duration." For its contradictory, "Time is not duration" or "It is false that time is duration," is perfectly non-absurd—and necessarily true. It is necessarily true by the very meaning of "time" and "duration." The statement predicates of time the attributes of duration; i.e., properties some of which, at least, are incompatible with the properties it possesses as time. Or, in semantic terms, at least part of what "duration" means₁ or signifies is incompatible with what "time" signifies. Now consider "Friday (the fifth day of the week) is in bed." This statement is absurd₁; but its contradictory, "Friday is not in bed" or "Friday is out of bed" is also absurd, unlike "Time is not duration." Hence "Friday is in bed" cannot be self-contradictory; since if it were its own contradictory would be necessarily true and hence non-absurd. For no absurd statement can possibly be true (or false), let alone necessarily true; and vice versa.

The upshot is that some self-contradictory statements are also (called) absurd₁; and some absurd₁ statements involve a self-contradiction. Consequently it is false that "A sentence which is a category mistake cannot get to be contradictory."¹⁸ It is simply false that (as Fred Sommers holds) if a self-contradictory statement commits a type-mistake "it would make no sense to call it an inconsistent or self-contradictory sentence."¹⁸ But it is true that any statement of the form "A is B and not B" is not absurd₁; and that we cannot say that "inconsistent sentences violate two rules, type rules and rules governing the affirmation and denial of predicates. It can in fact be shown that all expression-pairs of the form (X, not X) are U in sense value."¹⁹

We said before that absurd₁ statements may be vague in some degree or other; or, put the other way round, that vague statements may also be absurd₁. But this is only possible if they are slightly or moderately vague; and even then it is frequently difficult to determine whether or not a vague statement is also absurd₁. If a statement is extremely vague we cannot at all determine whether any absurdity is involved; just as we cannot determine what an extremely hazy snapshot is (intended to be) the picture of. Indeed, we cannot even properly raise the question whether or not these statements are absurd₁; and even,

whether or not they are self-contradictory. We cannot also meaningfully affirm, or deny, that a statement can in principle be both extremely vague and absurd₁. Hence it would be improper to speak of extremely vague statements in general as committing or not committing any type-mistakes.

Sense (2). We now turn to the second sense of "absurd" distinguished above. It is clear that a statement (or a set of statements) which expresses a belief, supposal, theory, or view that is regarded as absurd in the present sense (hereafter referred to as "absurd₂") by someone or other at a given time, will have as much—and as coherent—a meaning₁ as any non-absurd statement or proposition, in both senses (1) and (2) of "absurd." For example, the statement "Unsupported bodies fall toward the center of the earth because spirits pull them down" would be nowadays generally regarded as absurd as an alleged explanation of what we call the pull of gravitation. But as a statement it has a perfectly coherent meaning₁ (I am assuming, for purposes of illustration, that the word "spirit," in the sense intended here, has a meaning₁). It is not, with respect to its meaning₁, like "The $\sqrt{2}$ is asleep" or "My toothache is triangular." Its absurdity is not due to any peculiarity or irregularity in its meaning₁ as such. It is regarded as absurd by someone at a given time—note the relativity involved—by virtue of the character of the actual or possible state(s) of affairs to which it refers or about which it states something. More precisely, it is frequently regarded as absurd because of its alleged violent conflict with (what the judge considers to be) a great many empirical facts, or a whole set of well-established scientific theories or laws. (If the statement is believed to conflict with the "laws of reason or logic"—e.g., if it is self-contradictory—it will be absurd in the *third* sense of the word, which is closely related to its present sense.) When a person regards a particular statement as absurd, he generally does not merely regard it as false, he regards it as contrary to the whole tenor of his fundamental convictions concerning the nature of the world; in drastic conflict with the theories or principles he accepts. The same applies, *mutatis mutandis*, to supposals, beliefs, theories, or views.²⁰

This analysis of absurdity₂ does not apply to

¹⁸ *Ibid.*, p. 181.

¹⁹ *Ibid.* The proofs are given on p. 181 (n. 1) and p. 182.

²⁰ The word "absurd₂" is, however, sometimes used as a hyperbole in relation to statements, propositions, beliefs, etc.: usually for humorous purposes. There it is used merely as an exaggeration for "false." In this use, the word is frequently applied to *trivial* supposals, beliefs, and so on.

plans, schemes, or courses of action which are or may be regarded as absurd by someone at a given time. However, it would apply to them to the extent to which certain beliefs or assumptions may prompt a particular plan, scheme, or course of action in question, or determine its actual character to a greater or lesser extent. The analysis of the notion of absurdity as it arises in relation to schemes, plans, and courses of action does not concern us here. Let me say merely that, as with (alleged) absurd beliefs, statements, etc., an (alleged) absurd plan, scheme, etc., is one which is (regarded as) *unreasonable* or *irrational*—but in a different sense of these words than the sense involved in relation to the former.

It is noteworthy that absurdity, both in relation to beliefs, statements, etc., and in relation to plans, schemes, or courses of action, admits of degrees. For instance, sometimes we say: "*A*'s (business, military, educational, social, etc.) scheme is less (more) absurd than *B*'s (business, military, etc.) scheme"; "This theory (or view) is more (less) absurd than the one it is intended to replace"; or "The belief in the possibility of minds existing without brains is as absurd as the belief, in the 20th century, in ghosts and spirits." However, the degree of absurdity₂ of a particular belief (statement, supposal, theory, or view) *A* according to the particular judge *P*, either relative to another absurd₂ belief (statement, etc.) *B*, or to non-absurd beliefs (statements, etc.), has nothing to do with *A*'s or *B*'s meaning *as such*. The degree of absurdity₂ ordinarily assigned to *A* is determined, very roughly, by how flagrantly or glaringly it is judged by *P* to go against those of his convictions that are relevant to *A*. That is, it is determined by *P*'s rough and ready estimate of how much of his beliefs or convictions, in the relevant sphere, he must abandon or modify if he were to regard *A* as true, or even as possibly true. It is clear from this that the degree of absurdity₂ assigned to *A* may vary with different judges at a given time or at different times; also, that at time *T*, *A* may be regarded as absurd₂ by a particular person *P* and as perfectly non-absurd₂ by another person *Q*. The ordinary concepts of absurdity₂ and non-absurdity₂ are, however, open-textured: there is no hard-and-fast demarcation-line between absurd₂ and non-absurd₂ beliefs, statements, and so on. (Much the same holds, *mutatis mutandis*, with plans, schemes, or courses of action that are or may be said to be absurd₄ by some judge *P* at time *T*.)

Finally, it is clear that *A* may be vague, as well as absurd₁ (or absurd₁) according to a particular person or group of persons *P* at time *T*. But this will be true only if *A* is slightly or moderately vague; and so can be meaningfully said to be absurd or non-absurd. For the extreme indefiniteness of the content of an extremely vague belief, statement, etc., makes it in principle impossible for it to be or not to be in conflict with any belief, statement, etc., that *P* (or anyone else for that matter) regards, or may regard, as *well-established* or *true*.

The same is true, *mutatis mutandis*, in the case of a plan, scheme, or course of action *C* that a particular person or group of persons *P* may, at time *T*, regard as absurd₄. For the extreme indefiniteness of the content of an extremely vague plan, scheme, or course of action makes it in principle impossible for it to be or not to be in conflict with any plan, scheme, or course of action that *P* (or anyone else for that matter) regards, or may regard, as *reasonable* or *rational*.

V

It is instructive in a discussion of semantic vagueness—vagueness arising in relation to language—such as ours to make a comparison between it, in both our uses of the word (what we shall say, with one exception, applies indiscriminately to both), and *noise*, in a certain technical use of this word. I am referring to the concept of noise as it arises in relation to communication, and the theory of information. In this sense—

"Noise" refers to any disturbances or interference, apart from the wanted signal or message. This factor is always present to some degree, in every type of communication link whether electrical or not. "Noise" is the ultimate limiter of communication. The physical form of it which has most been studied is the *random* motion of the electrons in the various conductors and tubes in electrical apparatus.²¹

The following are some of the points of comparison between the two:

(a) The definition of "noise" given above, and the rest of the quoted passage, show us the basic analogy or similarity between it and vagueness. They both constitute an interference with, or are limiters of communication; and this gives them considerable practical importance. But the concept of "noise" is broader than that of (semantic)

²¹ E. Colin Cherry, "The Communication of Information," *American Scientist*, vol. 40 (1952), p. 646. Italics in original.

vagueness: "noise" interferes not only with communication, as does vagueness, but with all sorts of other activities or states of affairs; e.g., experiments or other physical situations involving the use of electric circuits, where measurements are to be made.²² On the other hand it is, like vagueness, only one major cause of loss of information in its particular domain. Ambiguity is another major limiter of communication by means of everyday language (also, the semi-technical language of many philosophers but rarely if at all the language of the advanced sciences). In the case of physical or electrical systems, information may be lost through errors in coding the message for transmission. This occurs before the information is transformed into a physical or electrical impulse or signal; hence before "noise" is generated in the circuits, etc.

(b) Although vagueness is always a defect in language with respect to the transmission of *information* in the sense of propositions, of something that is either true or false, just as "noise," as defined, is (necessarily) a limitation on communication,²³ it is far from a defect when used in literary compositions for literary aesthetic effect. Similarly it is not a defect when used in everyday situations, where the speaker or writer aims to evoke certain attitudes, feelings, or emotions in the reader or hearer. (Non-semantic forms of vagueness are also used in a similar way in painting and sculpture.) On the other hand it is sometimes deliberately used by lawyers and politicians as well as the man in the street to mislead the reader or hearer, e.g., by giving him a false impression of, or by leaving him in doubt as to what is, the real state of affairs.

(c) One difference between vagueness and "noise" is given by the definition of "noise" itself. "Noise" is an interference "*apart from* the wanted signal or message." The interference is due to motions or vibrations (e.g., sound waves) that are distinct from the signals "containing" the transmitted information. Vagueness, by contrast, is a "feature" of the verbal expressions that are themselves part or the whole of the vehicle for the transmission of the information in question. It is not, like "noise," a source of interference external to the vehicle of communication.

(d) A further difference between vagueness and "noise" pertains to the notion of expenditure of energy. The production of "noise" in any system

involves the expenditure of energy—energy that could be, partly (in practice) or wholly (ideally) used to transmit or help transmit the intended information. The notion of expenditure of energy in the *sending out* of information has no counterpart in the case of vagueness. For though physical and mental energy is expended in the transmission of a (vague) passage orally or in written form—as well as, of course, in thinking out the information (ideas, propositions) to be transmitted and their exact linguistic expression—there is no evidence, as far as I know, that more energy is expended in the transmission of vague language than non-vague language (the latter being the analog of transmission of physical or electrical signals in the absence of "noise"). What limited and pre-scientific "evidence" we possess tends to show, I think, that the expression (not the transmission) of ideas in vague language, *in our second but probably not our first* sense of "vague," involves less effort, as we ordinarily say, than the expression of the same ideas (assuming that they themselves are not vague) in precise language. However, the notion of expenditure of energy in the generation of "noise" in physical or electrical systems has at least a rough analog in the case of the *reception* of information expressed in vague language. The receiver (hearer or reader) must make a special mental and, sometimes, also physical effort in order to try to get the intended information. This is particularly true when the language in which the information is expressed is very vague.

(e) Finally for our present purposes, let us note one important difference between vagueness and "noise" which pertains to meaning. The vagueness of a vague expression is a semantic matter, is due to its possessing a meaning which fails to convey adequately the idea or ideas it is, or may be, intended to convey. "Noise" as defined, on the other hand, logically has nothing to do with meaning.

I might add by way of conclusion that the problem of loss of information due to the vagueness of the symbols in which it is expressed arises in actual communication by means of physical or electrical devices of communication such as telegraphs or telephones, as well as in the case of reading a book, say, written in ordinary English. I might also add that the problem of vagueness, which, as far as I know, has not yet been con-

²² See Leon Brillouin, *Science and Information Theory*, second edition (New York, 1962), pp. 125-126.

²³ Throughout, I use "communication" in the sense of "transmission of information" as ordinarily (and as above) characterized. I do not include in this notion the emotive and dynamic aspects of everyday language.

sidered in relation to the *theory* of information,²⁴ must be tackled, sooner or later, by writers on the subject. For instance, it must be considered in attempts to devise methods for the measurement of the amount of information contained in a given sentence that is to be transmitted. (Also, in this connection, a quantitative measure of degrees of (semantic) vagueness must be devised. This has not really been done. For instance, Max Black's proposals to give a quantitative account of degrees of vagueness in the paper alluded to earlier are really proposals to give open-texture, with which Black confuses vagueness, a quantitative measure.) Second, this problem must be dealt with in the construction of artificial languages for the transmission of (as much as possible of) the mass of information normally expressed in ordinary language. The expressions occurring in such languages must be, clearly, free from all vagueness.

VI

It will be recalled that I said in Section I that vague statements or propositions—and we must now add: *those whose vagueness is a result of their containing one or more vague expressions*—cannot be properly said to be either true or false, or either analytic or synthetic. I shall end this paper with some remarks about vague *concepts*, which are closely related to this thesis. Let us take as our example G. E. Moore's concept of sense-data, whose vagueness has been, I think, definitely shown by the many unsuccessful attempts of his critics to "locate" sense-data in our perception of physical objects; and by Moore's equally unsuccessful attempts to give clear directions for "locating" them. Despite the vagueness of this concept, *some* (but not all) of the statements which Moore makes about sense-data, or which we can make in the light of his characterization of them, are precise, not vague, and therefore *are* either true or false, analytic or synthetic. The majority of these non-vague statements, however, assert what sense-data are not; they state or are in other ways about what distinguishes them from other things, e.g., material objects. Thus on the basis of Moore's delineation of the concept of sense-data, the statements "The sense-data of a hand are not identical with the hand of which they are the sense-data," and "The concept of a hand is not logically analyzable, without residue, into the concepts of the sense-data of the hand," are analytically true. (Moore himself

would definitely hold, I think, that they are empirically true.) At the same time, we can form at least one positive statement about sense-data which, if interpreted in a certain way, will have precise meaning, and can be regarded as either true or false, as analytic or synthetic. Thus Moore held that we perceive sense-data in a primary sense of "perceive." We can therefore plausibly regard the statement "The sense-data of an object *x* are perceivable by any normal human being under certain physical, physiological, and (perhaps) psychological conditions" as *analytically* true within the framework of Moore's sense-datum "theory" conceived as an "alternative language," as proposing a set of concepts which reciprocally determine or delimit one another. On the other hand, if this statement is regarded—as it would be by Moore himself—as purporting to be an empirical statement, we cannot in principle properly assert that it is either true or false; since as we said before, Moore does not give us any clear directions for locating sense-data in our perception, as we say, of physical objects. As a consequence, it will be impossible in principle either to confirm or to disconfirm this statement.

The "boundaries" between Moore's concept of sense-data and other concepts are fairly well drawn by Moore, at least in some directions. But positively speaking the concept is far from well-defined: what lies or is supposed to lie "within" its "boundaries" is far from definite or clear. And it is this vagueness in the content of the concept that is the main source of its vagueness as a whole. Indeed, if its contents were well-defined and clear, a good part of the indefiniteness in its boundaries with other concepts would have vanished. If what it is were clear, what it is not would also have been clear. (Compare this to the situation in the case of marginal indeterminacy or open-texture.)

The Moorean concept of sense-data is not extremely vague. For extremely vague concepts are such that no (or almost no) negative—and not merely no positive—precise statements can be formed with their help. This means that none or only extremely few of the statements formed with their help can possibly be either true or false, analytic or synthetic. Other philosophical concepts roughly of the same order of vagueness as Moore's concept of sense-data are, I think, the Platonic and the (philosophical) Medieval concept of immortality of the soul, as well as some philosophical concepts of the nature of the soul itself.

²⁴ See Brillouin, *op. cit.*, chapters 18 and 20, and *passim*. Cf. also the literature he cites.

Some traditional philosophical concepts occupy an intermediate position between Moore's concept of sense-data, say, and extremely vague concepts. Only negative statements that utilize one or more of these concepts can be properly said to be either true or false, analytic or synthetic. No precise positive statements can be formed with their help. Locke's concept of material substance and (by the nature of the case) Kant's concept of a thing-in-itself are, I think, outstanding examples.

The foregoing is intended to apply to vague concepts in our first kind of use or sense of "vague concept"; i.e., to concepts that do not possess a relatively-fixed and well-defined content determined by linguistic convention. Whether or not it applies also to vague concepts in our second kind of use or sense of this phrase—to concepts that lack a sufficiently specific content—is an interesting question. But we shall not attempt to answer it here.

The American University of Beirut

IV. RECENT WORK ON PRESOCRATIC PHILOSOPHY

G. B. KERFERD

THE following survey deals with the period from 1953 to 1962. The period has seen a steadily mounting interest in the Presocratics, with the volume of publications increasing year by year. One new periodical, *Phronesis*,¹ is wholly devoted to problems in Greek philosophy, and they bulk largely in the revived *Archiv für Geschichte der Philosophie*,² the new *Archiv für Begriffsgeschichte*,³ and the Italian *Rivista critica della Storia di Filosofia*. Publications are found in Japanese, Turkish, Serbian, Russian, and Czech as well as in other languages. Books proliferate and articles and discussions in classical and philosophical journals require many pages to list them each year.⁴ The problem here is to find significant lines of development among a mass of detailed studies. In what follows I ignore detailed discussions of particular fragments or problems unless they contribute markedly to the understanding of some wider question. I am also very conscious that I have not seen, let alone read, much that is likely to be of interest or importance—a few such items I mention and mark with an asterisk. But I am sure there are others as deserving of mention as those which I include.

One of the most important questions in the period has been the authority of Aristotle and Theophrastus for the understanding of the Presocratics. In the nineteenth century Diels showed that almost all the later statements in ancient authors about the Presocratics derived from Theophrastus' lost work, *The Opinions of the Physicists*, and he believed that in reconstructing the main contents

of this work he was recovering for us secure information about the doctrines of the Presocratics which could be used to supplement and interpret the scanty surviving fragments of their original works. In 1935 H. Cherniss had given expression to major doubts about what Aristotle tells us in connection with the Presocratics, showing that in many cases he restated or distorted their doctrines to make them fit within his own philosophic framework.⁵ In 1953 J. B. McDiarmid published an important article on "Theophrastus on the Presocratic Causes"⁶ in which he argued that although Theophrastus may seem to have had access to the original texts, in fact, and in general, he contented himself with reproducing and elaborating what he found in Aristotle. If this view were to be correct it would be nearly fatal to the traditional picture of the Presocratics built up by scholars over a hundred years. This extreme conclusion has in fact been drawn by some, e.g., by E. A. Havelock,⁷ and in the aphorism "we know less and less about the Presocratics every day." Even those who stop short of such extremes have been influenced by the thesis to a very considerable extent, and there is a strong sceptical tinge in much writing in the decade we are considering. But the other side has also been represented, above all by W. K. C. Guthrie, first rather cautiously in an article published in 1957 entitled "Aristotle as a historian of philosophy: some preliminaries,"⁸ and then resoundingly in his *History of Greek Philosophy*, vol. I,⁹ to which frequent reference will be made in what follows. In this

¹ Assen, Van Gorcum; the first number appeared in November, 1955.

² This was revived in 1960 after a lapse of 27 years.

³ Bonn, 1955 and subsequently. Some numbers are wholly devoted to Greek terms; others contain nothing relating to ancient philosophy.

⁴ The annual surveys in *L'Année Philologique* provide the fullest coverage known to me.

⁵ *Aristotle's Criticism of Presocratic Philosophy* (Baltimore, 1935), and his article, "Characteristics and Effects of Presocratic Philosophy," *Journal of the History of Ideas*, vol. 12 (1951), pp. 319-345.

⁶ *Harvard Studies in Classical Philology*, vol. 61 (1953), pp. 85-156.

⁷ *A Preface to Plato* (Oxford, 1963), pp. vi-vii. See also J. A. Philip, "The Fragments of the Presocratic Philosophers," *Phoenix*, vol. 10 (1956), pp. 116-123.

⁸ *Journal of Hellenic Studies*, vol. 77 (1957), pp. 35-41.

⁹ Cambridge, 1962. For other discussions see, e.g., Kahn, *Anaximander*, pp. 17-24 (n. 34 below); J. Kerschensteiner, "Der Bericht des Theophrast über Heraklit," *Hermes*, vol. 83 (1955), p. 385; and F. Solmsen, "Aristotle and Presocratic Cosmogony," *Harvard Studies in Classical Philology*, vol. 63 (1958), p. 277.

battle I would award the victory to Guthrie, but this is a matter on which debate will unquestionably continue. Most discussions of individual Presocratics, including those which are not directly influenced by Cherniss and McDiarmid, can be placed somewhere between the two poles of extreme scepticism and critical acceptance of the basis of the Aristotelian and Theophrastean tradition.

A second debate concerns the place to be assigned to religious and mythical elements in the thinking of the Presocratics. The traditional classicist view of the Greeks as wholly rational in outlook tended to present the Presocratics from Thales onward as marking a complete break with previous mythological thinking, and themselves as wholly scientific and rational in outlook. The idea that there was any break in continuity is now generally rejected, the change being marked in 1934 when Kranz in his edition of Diels' *Fragmente der Vorsokratiker*¹⁰ brought the section on "Early Cosmological Poetry" out of the Appendix and placed it before the treatment of Thales. This is reflected in the very large section dealing with "The Forerunners" at the commencement of Kirk and Raven's important work, *The Presocratic Philosophers*,¹¹ above all in the treatment of Pherecydes. Sometimes the emphasis upon continuity is coupled with a radical scepticism about the later, Aristotelian-style, or philosophized picture of the Presocratics, so as to stress the mythical and fanciful element in earlier thinkers and even in Parmenides and his successors. This is the tendency, I would say, in Kirk and Raven, and also in U. Hoelscher's articles, *Anaximander und die Anfänge der Philosophie*.¹² Guthrie in his *History*,

without rejecting the evidence of Aristotle and Theophrastus, carries through a highly successful reinterpretation of the Pythagoreans and Heraclitus which shows a profound understanding of the non-rational and traditional sides of their thinking, although he is inclined to retain perhaps rather too much of the traditional picture of the Milesians, and to underrate the mythical element in them. At other times it is the rational element in thinkers before Thales, and indeed in the myths themselves which have been emphasized.¹³ More and more what was happening is being seen as a gradual process, now often called *Entmythologisierung* in a slightly different sense from that in which that term was applied in Biblical scholarship, to mean a sort of historical process of change from mythical to rational formulations of a world-picture which changed only gradually in the process.¹⁴

A related but distinct problem is the correct philosophic assessment of the rational element unquestionably present in the thought of individual Presocratics whatever its extent or limitations. Some have been concerned to detect the beginnings of metaphysical as distinct from physical theorizing, and have traced this back as far as Anaximander¹⁵ without much real plausibility. Parmenides is still generally considered the first of these,¹⁶ but if the doctrine that a thinker may be quite unconscious of what he is in fact doing is added to the view of the Presocratics as representing a continuous development from earlier thought, then anticipations may be traced back almost indefinitely. Others have been concerned with the origins of science. Guthrie sees the beginnings of science in the naturalism of the Ionians¹⁷ and in their search for an underlying impersonal order.¹⁸

¹⁰ Fifth edition. The sixth edition, 1951-52, is a reprint of the fifth with some 40 pages of extra notes, but subsequent "editions" down to and including the tenth are unchanged reprints of the sixth.

¹¹ Cambridge, 1957.

¹² *Hermes*, vol. 81 (1953), pp. 257-277, 385-418.

¹³ See J. P. Vernant, *Les origines de la pensée grecque* (Paris, 1962); H. Schwabl s.v. "Weltschöpfung" in Pauly-Wissowa, *Realencyclopädie* Supp. IX (1962), pp. 1433-1582, especially pp. 1513-1537; A. Rivier, "Pensée archaïque et philosophie présocratique," *Revue de Théologie et de Philosophie*, vol. 3 (1953), pp. 93-107; and "Sur le rationalisme des premiers philosophes grecs," *ibid.*, vol. 5 (1955), pp. 1-15. The treatment by J. Chevalier, *Histoire de la Philosophie*, vol. 1 (Paris, 1955) is much too traditional here.

¹⁴ Similar use is made of the term "désacralisation," e.g., in J. P. Vernant (n. 13 above). No particular English term seems yet to have been brought into common use.

¹⁵ So in an extreme way P. Seligman, *The Apèiron of Anaximander, a Study in the Origin and Function of Metaphysical Ideas* (London, 1962).

¹⁶ So in quite different ways D. Grey, "Parmenides and the Invention of Metaphysics," *Bulletin of the Institute of Classical Studies* (University of London), vol. 7 (1960), pp. 67-68, and G. E. L. Owen, "Eleatic Questions," *Classical Quarterly*, vol. 10 (1960), pp. 84-102.

¹⁷ See his *In the Beginning, Some Greek Views on the Origins of Life and the Early State of Man* (London, 1957), chap. 6.

¹⁸ *History*, vol. 1, pp. 26-29. Relevant here is O. Gigon, "Die Theologie der Vorsokratiker," *Fondation Hardt, Entretiens Tome I, La Notion du Divin* (Geneva, 1954), and J. Kerschensztein, *Kosmos, Quellenkritische Untersuchungen zu den Vorsokratikern* (Munich, 1962).

But if such a wide view is taken of science then surely it was already present both in Greek and non-Greek thought to the extent that *some* non-divine explanations of phenomena were known, and an underlying impersonal order is presented in many early myths which are so fantastic that we would not normally wish to call them scientific. Another approach asks how far the Presocratics used observation or experiment, and how far their thinking was *a priori*. Popper in a much discussed address¹⁹ has maintained that the Presocratics worked from theories and speculations and not from observations or experiments. He has been answered by Kirk²⁰ who claims convincingly that "imaginative" observation (but not experiment) played a large part in many theories, but is on weaker ground in suggesting that if Thales used mythical conceptions this merely means that he is using the stored observations of men before him.

The problem of "philosophic" interpretations of the Presocratics remains acute, quite apart from the question of distortion by Aristotle or Theophrastus. The way in which philosophers like Hegel, Nietzsche, and the neo-Kantians falsified the history of Greek philosophy by finding in it too much of their own theories is well known, but is seldom taken as an object lesson. If Aristotle read into the Presocratics assumptions that were of his own making we find it just as easy to do the same with our own philosophic presuppositions, although we protest with indignation at attempts to do this with presuppositions which we do not ourselves hold. Philosophical presuppositions are largely

absent in Kirk and Raven's book, and in Guthrie's *History*.²¹ They present the Presocratics as cosmologists and scientists, and on occasion as theologians and mystics, but generally speaking not as philosophers in any modern sense of the term. As a result many feel that there is something that they have missed.²² Existentialist interpretations tend to seem obviously wrong both in method and in resulting interpretation to most members of the English speaking world.²³ The same tends to be the case with the idealist and neo-Hegelian trend of some Italian scholarship, above all the numerous writings and commentaries of M. Untersteiner.²⁴ An analytical and logical approach to problems tends to seem natural and proper to those brought up in the English analytical tradition, and there is a danger that an interest in such questions may be identified over-readily, or to an exaggerated extent, in the Presocratics, and that this should not even be suspected by those who are doing it.²⁵ The only remedy is eternal vigilance and the determined use of the whole apparatus of scholarship applied to all the evidence before any interpretation is propounded—mere intuitive accounts can no longer be accepted as sufficient.

Certain other themes of hardly less importance must be mentioned more briefly. The attitude of the Presocratics toward the idea of subjective existence is discussed in R. Mondolfo's large work, first published in Spanish in 1955.²⁶ Ethical thought is set in a new and wider context in an important book by A. W. H. Adkins, which falls partly outside the scope of this survey.²⁷ The history of

¹⁹ "Back to the Presocratics," *Proceedings of the Aristotelian Society*, vol. 54 (1958-59), pp. 1-24, reprinted as chap. 5 in his *Conjectures and Refutations* (London, 1963). See also his *Logic of Scientific Discovery* English ed. (London, 1958). For Popper there is nothing unscientific in such proceedings; rather the reverse is the case.

²⁰ "Popper on Science and the Presocratics," *Mind*, vol. 69 (1960), pp. 318-339, and "Sense and Common-sense in the Development of Greek Philosophy," *Journal of Hellenic Studies*, vol. 81 (1961), pp. 105-117. Popper has replied in an appendix to chap. 5 of his *Conjectures and Refutations* (n. 19 above).

²¹ Notes 9 and 11 above.

²² Kirk and Raven even exclude the sophists, for the unintentionally revealing reason that "their positive philosophical contribution, often exaggerated, lay mainly in the fields of epistemology and semantics." (Pref. p. vii).

²³ Heidegger's discussion *Der Spruch des Anaximanders*, subsequently published in his *Holzwege* (Frankfurt, 1950), and discussed by C. Ramnoux in her article "Sur quelques interprétations modernes d'Anaximandre," *Revue Métaphysique et Morale*, vol. 59 (1954), pp. 233-252. This outlook seems to have influenced her subsequent *Héraclite ou l'homme entre les choses et les mots* (Paris, 1959), and, despite disclaimers, L. Winterhalder, *Das Wort Heraklits* (Zurich, 1962). For Parmenides see J. Beaufret's *Introduction to his Le Poème de Parménide* (Paris, 1955), and see also K. Jaspers, *Die Grossen Philosophen*, Bd. I (Munich, 1957), pp. 619-655. J. Bollack, "Die Metaphysik Empedokles als Entfaltung des Seins," *Philologus*, vol. 101 (1957), pp. 30-54 is perhaps rather Kantian than existentialist in character.

²⁴ *Senofane, Testimonianze e Frammenti* (1955), *Parmenide* (1958), *Zenone* (1963), *I sofisti*, Fasc. i-iv (1949-1962), all in the series *Biblioteca di Studi Superiori* published at Florence, his separate book *I Sofisti* (Turin, 1949), English tr. (Oxford, 1954), and numerous articles listed in his books.

²⁵ A controversial example might be Owen's article (n. 16 above) on which see below.

²⁶ *La Comprension del sujeto humano en la cultura antigua* (Buenos Aires, 1955); *La Comprensione del soggetto umano nell'antichità classica* (Florence, 1958).

²⁷ *Merit and Responsibility, a Study in Greek Values* (Oxford, 1960).

Presocratic thought is analyzed in Marxist terms in George Thomson's *Studies in Ancient Greek Society*, vol. II,²⁸ and a number of key terms and concepts have been the subject of important studies which must be taken into account in any future attempts to make general statements about the course of development of Presocratic thought.²⁹

I turn now to detailed studies of one or more of the individual Presocratics.

1. **THALES.** Kirk, following in this U. Hoelscher,³⁰ emphasizes the probable near-Eastern (? Egyptian) origin of Thales's thinking, and doubts whether he really held that the earth *consists of* water, even if he did think that it originated from water, while Guthrie³¹ holds that "at the conscious level, he had made a deliberate break with mythology and was seeking a rational account." These will no doubt for long continue to be the two main attitudes taken toward Thales. In either case it is probable that his importance as the founder of Greek philosophical thinking has been exaggerated, but few will wish to go quite as far as D. R. Dicks in dismantling the whole tradition of his achievements in mathematics.³² I find nothing new in the survey by E. Stamatis³³ who is more concerned with Thales as one of the traditional sages than with his physical doctrines.

2. **ANAXIMANDER.** Here there have been a number of really important discussions. As a result above all of C. Kahn's book³⁴ we have a much

clearer picture of the nature of the doxographic tradition, we can pose questions more precisely, and it is possible that we can even answer some of them more confidently. To take details first: it is becoming doubtful whether he thought of his *Apeiron* as an *Arche* or first principle in Aristotelian terms, and he probably did not speak of it as an *Arche* at all.³⁵ His *Apeiron* may well have been spatially unbounded rather than qualitatively indeterminate,³⁶ and its god-like "controlling" influence is now at last fully recognized.³⁷ Whether he ever envisaged an infinite series of worlds either successively or simultaneously becomes increasingly doubtful, and it is fairly clear that for him at least the term *Kosmos* did not mean world or universe, but rather any one of a number of subordinate areas within the whole.³⁸ The tradition as to how these *Kosmoi* developed out of the *Apeiron* is incomplete and confused in detail, but the main features of the process can be reconstructed with considerable plausibility.³⁹ Most important of all we seem to be reaching a clearer understanding of the famous fragment—the things that pay the penalty to one another must in some way or other be the opposites, and they do so by passing into one another and not back into the *Apeiron*.⁴⁰ All this involves a fairly large change from traditional pictures of Anaximander. Though controversy will continue, the Nietzschean interpretation positing the wickedness of the existence of ordinary things must now finally be banished. Anaximander

²⁸ *The First Philosophers* (London, 1955).

²⁹ **Kosmos:** J. Kerschenshteiner (n. 18 above) and W. Kranz, *Archiv für Begriffsgeschichte*, vol. 2 (1955), pp. 1–262. **Aition:** H. Boeder, *Revue des Sciences Philosophiques*, vol. 40 (1956), pp. 421–442. **Arche:** Lumpe, *Archiv für Begriffsgeschichte*, vol. 1 (1955), pp. 104–116. **Elementum:** W. Burkert, *Philologus*, vol. 103 (1959), pp. 167–197; Lumpe, *Reallexicon für Ant. und Christ.*, vol. 4 (1959), pp. 1073–1100, and *Archiv für Begriffsgeschichte*, vol. 7 (1962), pp. 285–293. **Logos** and **Altheia:** H. Boeder, *Archiv für Begriffsgeschichte*, vol. 4 (1959), pp. 82–112. Here may be mentioned also R. Joly, *Le Thème philosophique des genres de vie dans l'antiquité classique* (Brussels, 1956).

³⁰ Kirk & Raven, pp. 74–98; cf. Hoelscher in *Hermes*, vol. 81 (1953), pp. 385–391.

³¹ *History*, vol. 1, p. 62.

³² *Classical Quarterly*, vol. 9 (1959), pp. 294–309. A more positive view is that of A. Wasserstein, *Journal of Hellenic Studies*, vol. 75 (1955), pp. 114–116; and vol. 76 (1956), p. 105.

³³ *Das Altertum*, vol. 6 (1960), pp. 93–103.

³⁴ *Anaximander and the Origins of Greek Cosmology* (New York, 1960).

³⁵ So Kirk in *Classical Quarterly*, N.S. vol. 5 (1955), pp. 21–23, and Kirk & Raven, pp. 107–108; McDiarmid in *Harvard Studies in Classical Philology*, vol. 61 (1953), pp. 138–140; Guthrie, *History*, vol. 1, p. 77 (with qualifications). The contrary view in Kahn, *Anaximander*, pp. 29–32, Seligman, *Anaximander*, pp. 19–20, 26–28, and Vlastos in *Gnomon*, vol. 27 (1955), p. 74, n. 2.

³⁶ Kahn, *Anaximander*, appendix II, *Classen in Hermes*, vol. 90 (1962), pp. 159–166. I find nothing plausible in the argument of B. Wismiewski, *Revue des Études Grecques*, vol. 70 (1957) that the *Apeiron* was spatially finite. That it was qualitatively indeterminate is still maintained by Kirk and by Guthrie.

³⁷ So Kirk, Guthrie, Seligman, C. J. Classen, and also F. Solmsen in *Archiv für Geschichte der Philosophie*, vol. 44 (1962), pp. 109–131.

³⁸ Kirk, Kahn, and Kerschenshteiner, *Kosmos* (n. 18 above). Guthrie holds to the Cornford theory of innumerable worlds.

³⁹ For a full discussion with diagrams see N. Rescher, "Cosmic Evolution in Anaximander," *Studium Generale*, vol. 11 (1958), pp. 718–731.

⁴⁰ Kahn (see my review in *Classical Review*, vol. 12 (1962), pp. 34–35), Broecker in *Hermes*, vol. 84 (1956), pp. 382–384. The truth was seen earlier by Vlastos, *Harvard Studies in Classical Philology*, vol. 42 (1947), pp. 168 ff. The wrong view in Hoelscher, *Hermes*, vol. 81 (1953), pp. 257 ff.

appears as primarily interested in developing a highly individual cosmology—whether he even thought of the question of a primary substance as such becomes increasingly doubtful. His philosophic importance probably lay rather in the implications of what he said, and these implications were probably much clearer to Aristotle than to himself. It is even possible that he was quite unaware that he was raising any questions that were not purely physical in character.

3. ANAXIMENES. Here there are only points of detail to report. Kirk (in Kirk & Raven) stresses the evidence that Anaximenes thought of the world as a living creature, while Guthrie is more concerned with the reasons which led him to regard Air as a primary source, but the difference is one of emphasis only. Nor is there much of importance to report about later Ionian physicists, although Diogenes of Apollonia has been the subject of some rather far-reaching claims by J. Zafiropulo.⁴¹

4. PYTHAGORAS AND THE PYTHAGOREANS. Here even more than for Anaximander everything depends upon the assessment of the sources. There are some signs that a period of extreme scepticism concerning the possibility of knowledge about, or even the existence of, important Pythagorean doctrines in the fifth century may be passing,⁴² but no generally agreed picture is yet in sight. However, the existence of a developed Pythagorean cosmology in the fifth century is beginning to seem more likely than it once did, and there is a growing readiness to make at least cautious use of the fragments of Philolaus in this connection.⁴³ There is also perhaps a better understanding of the way in which religious mysticism and fantasy could combine with a rationally elaborated cosmology,

whether in the thought of Pythagoras himself, or of his successors.⁴⁴

Assuming that there was a fully-developed Pythagorean cosmology at least by the middle of the fifth century, and that it is to this that Aristotle refers, the following have been the more important points discussed. It is now generally supposed, I think, that the system involved an ultimate dualism based upon the Limit and the Unlimited, and so should be contrasted with the monistic tendencies of the Milesian systems. This involves the rejection of F. M. Cornford's thesis that there was an ultimate One behind all else, although the importance of the One within the system is not in question. How then is this One, which may well have been regarded as divine, to be related to the Limit and the Unlimited? Probably it should simply be identified with the One which was one of the opposites and stood in the column headed by the Limit.⁴⁵ Whether either Parmenides or Zeno was attacking Pythagorean doctrines as such has begun to seem more and more doubtful,⁴⁶ and as a result it is also doubtful whether any modifications were made in Pythagorean doctrine to meet such attacks.⁴⁷ One possibility is that while Pythagorean physical doctrines were fully elaborated at an early date, they were not generally known to other thinkers until well into the second half of the fifth century. On the method by which things come into existence out of numbers, Guthrie⁴⁸ has argued against the common view that what Aristotle says in the *Metaphysics* involves three mutually inconsistent accounts—the Pythagoreans may well have maintained that things *are* numbers, that the principles of numbers are the principles of things, and that things *imitate* numbers, just as in effect Plato could maintain all three when considering

⁴¹ *Diogenes d'Apollonie* (Paris, 1956) with my review in *Classical Review*, vol. 8 (1958), pp. 185–186. I do not deal separately with Zafiropulo's other writings, notably *Empédocle d'Agrigente* (Paris, 1953), *Vox Zenonis* (Paris, 1958), where he gives expression to his "animistic" interpretation of the Presocratics according to which all, including Parmenides and Heraclitus, offered a view of reality as combining the spiritual and the material in a way never subsequently possible. See also chap. iv of S. Zeppi's book (n. 110 below).

⁴² H. Thesleff, *An Introduction to the Pythagorean Writings of the Hellenistic Period* (Abo, 1961), p. 45, is premature in claiming that this movement of thought is universal, although his own discussion points strongly in this direction.

⁴³ So holds Guthrie, vol. 1, pp. 329–333. On the other hand Raven is sceptical (Kirk & Raven, pp. 308–311), and so is W. Burkert, "Hellenistische Pseudo-pythagorica," *Philologus*, vol. 105 (1961), pp. 16–43, "Plato oder Pythagoras?" *Hermes*, vol. 88 (1960), pp. 159–177, and in his large book *Weisheit und Wissenschaft, Studien zu Pythagoras, Philolaus und Platon* (Nürnberg, 1962). In Italy the positive approach initiated by Mondolfo is continued in the editions of the fragments by M. Timpanaro Cardini, *I Pitagorici* Fasc. I (Florence, 1958), Fasc. II, 1962, and A. Maddalena, *I Pitagorici* (Bari, 1954).

⁴⁴ See, e.g., J. S. Morrison, "Pythagoras of Samos," *Classical Quarterly*, vol. 50 (1956), pp. 135–156; "The Origins of Plato's Philosopher Statesman," *ibid.*, vol. 52 (1958), pp. 198–218.

⁴⁵ So Guthrie, vol. 1, pp. 243–244. *Contra*, e.g., Burkert, *Weisheit* (n. 43 above), p. 33, n. 110. The most likely alternative is that the One is a compound of the Limit and Unlimited.

⁴⁶ See n. 70 and n. 73 below.

⁴⁷ As against Raven (in Kirk & Raven, p. 236).

⁴⁸ *History*, vol. 1, pp. 229–231.

the relationship between Forms and Particulars. This may be the correct explanation of the alleged Number-Atomism of the Pythagoreans. Two further points are of importance in Guthrie's treatment of Pythagoreanism—his warning of the weakness of the evidence that Pythagoras discovered the numerical basis of the intervals in the musical scale, and derived from it the theory that things are numbers,⁴⁹ and his argument that the doctrine of the soul as *Harmonia* is not inconsistent with a transmigratory soul, provided the harmony is one of the soul's own parts, and not as for Plato in the *Phaedo* a harmony of the body.⁵⁰

5. XENOPHANES. The only full-scale study in the decade is that of M. Untersteiner in his book published in 1955.⁵¹ This proposes a series of novelties, but none of them is soundly based. Above all his use of the pseudo-Aristotelian treatise *De Melisso Xenophane Gorgia* as a reliable source of information about Xenophanes' doctrines cannot be admitted, and while he may be correct in treating Xenophanes as a pantheist rather than a dualist, the way in which he does so does not carry conviction. While later sources were no doubt wrong in treating Xenophanes as standing at the beginning of Eleatic doctrine, it is becoming clearer how such a mistake was possible. In particular Guthrie,⁵² who treats him as identifying god and the world—and so as a pantheist—argues that while he excluded motion in the sense of locomotion from his world, and so had to that extent a static universe, this need not mean that he excluded change. In this way he provides an answer to Kirk's objection that god could not have been identified with the world, because if so he could not have been described as motionless.⁵³ The possible relation of this to Parmenides is surely clear.

6. HERACLITUS. Here the period in question is dominated by discussion of the series of interpreta-

tions propounded by Kirk,⁵⁴ which deal with virtually every problem raised by previous scholars, despite the fact that only a little more than half the surviving fragments, those bearing directly upon problems of the world as distinct from man, are included in his edition. The importance of Kirk's book was immediately recognized, but while some supposed that it would involve forthwith a decisive change in the way in which Heraclitus should be regarded, others have remained unconvinced by all the more radical changes proposed in the traditional picture.⁵⁵ Fundamental is Kirk's contention that the tradition about Heraclitus was radically distorted at a very early stage, and that what Plato and Aristotle say about his doctrines completely misrepresents his basic position. This distortion, it is claimed, was carried still further by Stoic interest in Heraclitus whom the Stoics wrongly supposed to have anticipated many of their distinctive contentions. So (1) the Flux doctrine, expressed by later writers in the formula "all things flow," and supposedly illustrated by the fragment "you cannot step twice into the same river" was not a doctrine held by Heraclitus according to Kirk, and the supposed fragment does not represent anything in Heraclitus that would support such a doctrine. Consequently if Plato's thinking rested as Aristotle indicated, and as the *Theaetetus* seems to show, upon a doctrine of the incessant flux of phenomenal objects, this did not come from Heraclitus. As against Kirk on these points, very strong arguments have been adduced by G. Vlastos and by Guthrie,⁵⁶ and there seems to be a growing conviction that in some sense Heraclitus did hold a doctrine of Flux. (2) The world is a harmony of opposites, Kirk claims, for Heraclitus in such a way that the tension between them enables things in the world of experience to be stable for a time. But Guthrie points out⁵⁷ that

⁴⁹ *Ibid.*, pp. 221-223.

⁵⁰ *Ibid.*, pp. 381-383.

⁵¹ See n. 24 above, and my review in *Gnomon*, vol. 29 (1957), pp. 127-131. I have not seen A. Farina, *Senofane di Colofone* (Naples, 1961). See also chaps. i-ii of S. Zepi's book, n. 110 below.

⁵² *History*, vol. 1, pp. 381-383.

⁵³ Kirk & Raven, p. 172.

⁵⁴ *Heraclitus, the Cosmic Fragments* (Cambridge, 1954), "Men and Opposites in Heraclitus," *Museum Helveticum*, vol. 14 (1957), pp. 155-163, "Ecpyrosis in Heraclitus, Some Comments," *Phronesis*, vol. 4 (1959), pp. 73-76, "Logos, Harmonie, Lutte, Dieu et feu," *Revue Philosophique*, vol. 147 (1957), pp. 289-299, "The extent of Heraclitus fr. 92D," *Annales de Filologia Classica* (Buenos Aires), vol. 7 (1959), pp. 5-12. Some of the ethical fragments are included in Kirk & Raven, pp. 204-214.

⁵⁵ Reactions fundamentally critical of Kirk's interpretations will be found in the review by Tate in *Classical Review*, vol. 6 (1956), pp. 20-22; the discussion by G. Vlastos in *American Journal of Philology*, vol. 76 (1955), pp. 337-368; R. Mondolfo, "Evidence of Plato and Aristotle Relating to the Ecpyrosis," *Phronesis*, vol. 3 (1958), pp. 75-82; I frammenti del fiume e il flusso universale in Eraclito, *Revista Critica di Storia di Filosofia*, vol. 15 (1960), pp. 3-13; Zeller-Mondolfo, *La Filosofia dei Greci I*, vol. 4, *Eraclito* (Florence, 1961); and Guthrie, *History*, vol. 1, pp. 409-492.

⁵⁶ Vlastos, *American Journal of Philology*, vol. 76 (1955), pp. 338-344; Guthrie, *History*, vol. 1, pp. 449-452 and 488-492.

⁵⁷ *History*, vol. 1, p. 452.

tension between opposites, and even identity of opposites, need not for Heraclitus have excluded cyclical change backwards and forwards as we would say between two poles. Here the metaphor, not used by Guthrie, of a tug of war where continuous tension does not exclude movement may be helpful. (3) Fire as a primary substance out of which the world is generated after the fashion of the Milesians, and into which it may be destroyed by *Ecpyrosis* is denied to Heraclitus by Kirk, although it is ascribed to him by Aristotle. As against this view held by Kirk, Vlastos has argued the case for relating Heraclitus to Anaximander and Anaximenes,⁵⁸ and Mondolfo⁵⁹ defends *Ecpyrosis* as well, although on this last point Guthrie accepts Kirk's view. Perhaps the analogy of the tug of war might be extended. If the cosmological fire which operates as an *arche* is not outside the world, but rather is the world, then in its aspect of strife or tension it never ceases to operate. From time to time one side in the tug of war collapses on the ground, and the immediate pattern of tension is destroyed. But the basic strife remains, the collapsed side rises to its feet and the world is renewed. (4) On the *Logos*, Kirk argues that this means an objective state of things, a sort of formula of arrangement common to all things, which on occasion could be treated as divine and identified with fire. He excludes, as Stoic only, interpretations which identify it with human reason and thinking. Against this Guthrie demurs⁶⁰ and insists that it must at least have included human thought for Heraclitus.

Other writings about Heraclitus can only be listed. Someone has spoken of the unwholesomeness of the attraction of Heraclitus for the modern world. Unwholesome or not, his supposed revolt against rational explanations has led to a whole stream of books. Many of these are not critical but intuitive or even mystical in their approach.

Kirk's work at the very least ought to have put all merely intuitive approaches out of court, although it may have led some to be over-impressed by the sheer *quantity* of scholarship which has been brought to the task of interpretation.⁶¹

7. THE ELEATICS. The most radical discussion of Parmenides in the period is due to Untersteiner.⁶² By emendation of the text he is able to argue that the question of the One was first raised by Melissus, and not by either Parmenides or Zeno. This involves the rejection of what Plato says in the *Sophist* and *Theaetetus*, as well as the evidence of Aristotle and Theophrastus and is completely unconvincing. More fruitful have been discussions about what the surviving lines of Parmenides may mean when not emended. First of all there is the problem of the subject of the verb "to be" in the poem when used without a subject directly expressed. There is something like agreement perhaps that the grammatical subject can only be a rather indefinite "it" or "a thing"—no grammatical subject is actually expressed, and statements such as "it is" occur too early in the poem for anything more specific to be understood.⁶³ But what of the logical subject? What is it that Parmenides is actually talking about? The view that it is the world or universe does not now seem to find so much favor as earlier.⁶⁴ G. E. L. Owen⁶⁵ argues that the subject cannot be "What is" as Diels and Cornford supposed, nor "The One" or "The One Being," and concludes that the subject is "what can be talked or thought about." He goes on to challenge the whole basis of what seems to have been Aristotle's view of Parmenides according to which Parmenides' thinking rested upon a confusion between the existential and predicative senses of the verb "to be," and he regards Parmenides' basic contention, as he interprets it, as a philosophic innovation comparable in importance with Descartes' *Cogito*.⁶⁶ He argues further that

⁵⁸ See n. 56 above at pp. 361-365.

⁵⁹ See n. 55 above.

⁶⁰ *History*, vol. 1, p. 426, n. 2.

⁶¹ A new work on Heraclitus on a vast scale (two volumes of which the first will run to 1,000 pages) is announced by M. Marcovitch for 1963. I have seen only the *Primera Parte* pp. 1-80, dealing with fr. 1 and published at Merida (Venezuela) in 1962. See n. 23 above for books by Ramnoux and Winterhalder. H. Quiring, *Heraclit, Worte tönen durch Jahrtausende* (Berlin, 1959) is a work by a physicist who has not read Kirk. P. Wheelwright, *Heraclitus* (Princeton, 1959) has been criticized for inaccuracy as well as deficiency in scholarship, but some of the criticisms merely assume alternative explanations have been proved when this is hardly the case. See also K. Axelos, *Heraclite et la Philosophie* (Paris, 1962).

⁶² See n. 24 above. He further discusses his view of Parmenides in an *Appendix* to his *Zenone* (1963).

⁶³ Untersteiner's attempt to argue that "The way" is the subject is hard to take seriously. J. H. M. Leonen, *Parmenides, Melissus, Gorgias* (Assen, 1959) would emend the text to yield an expressed subject "anything"—the emendation is unlikely, but the interpretation is likely. He goes on to maintain that Parmenides' Being in the Way of Truth is "necessary being" and is contrasted with contingent being in the Way of Truth, but this is hard to accept.

⁶⁴ But see L. Woodbury in *Harvard Studies in Classical Philology*, vol. 63 (1958), pp. 145-160.

⁶⁵ "Eleatic Questions," *Classical Quarterly*, vol. 10 (1960), pp. 84-102.

⁶⁶ The comparison with Descartes is already found, in a different sense, in Kirk & Raven, p. 266.

Parmenides did not claim any measure of truth or reliability for the cosmology in the "Way of Opinion," which merely represents the (erroneous) opinions of mortals, that his assumptions throughout do not derive from earlier cosmologies, and that he does not argue for the existence of a spherical universe. More than one of these contentions invites reply, but so far no full discussion has appeared. However, the belief that Parmenides' "Way of Opinion" had a higher status than Owen allows has much support,⁶⁷ and the details suggest that he gave his personal attention to their elaboration.⁶⁸ One may also ask is not the "Way of Opinion" at least something that can be thought *about* and spoken *of* although it clearly cannot be said to be about something that *is* in the sense in which this word is used in the earlier part of the poem.

According to Raven,⁶⁹ what Parmenides has done is "to take his own sphere of reality, the One, and fill it, quite illegitimately, with the sensible opposites of light and darkness." This remains a likely view. But less likely is the treatment of Parmenides' thought as involving a reaction against Pythagorean doctrine, or indeed any particular Presocratic.⁷⁰ It is more likely that he was concerned with the opinions of all mortals as he understood them. The older view that that which is thinks has been revived by E. D. Phillips.⁷¹ These and other questions will surely continue to

be debated. The unique position of Parmenides remains unquestioned, especially his claim to stand at the beginnings of the development of logic, although the temptation to make excessive claims has not always been resisted.⁷²

8. ZENO. The paradoxes of Zeno excite as much interest as ever, and have been the concern of modern thinkers who are not otherwise very much concerned with the Presocratics, as well as those who are deeply concerned with them. It is now quite widely supposed that he was not in his paradoxes attacking any particular previous thinker, but rather all those who would not or had not given their assent to the basic Eleatic contentions about Being.⁷³ The conclusive argument is perhaps the point brought out by G. E. L. Owen, namely, that the contradictions which he sets up only arise if two separate views are brought together. Discussion of particular paradoxes, primarily from the point of view of their logical solution, have been frequent.⁷⁴ Untersteiner in his edition of the fragments⁷⁵ discerns the two worlds which he found also in Parmenides—the world of being outside time and the world of sensible experience in time—and contends that Zeno set himself to reconcile the two by a new method of dialectic. The detailed analysis is necessarily a highly technical matter in the case of the Paradoxes, and there have been a number of important discussions which cannot be summarized here.⁷⁶

⁶⁷ See H. Schwabl, "Sein und Doxa bei Parmenides," *Wiener Studien*, vol. 66 (1953), pp. 50–75; W. R. Chalmers, "Parmenides and the Beliefs of Mortals," *Phronesis*, vol. 5 (1960), pp. 5–22; L. Woodbury, "Parmenides on Names," n. 64 above.

⁶⁸ See discussion by J. S. Morrison, "Parmenides and Er," *Journal of Hellenic Studies*, vol. 75 (1955), pp. 59–68.

⁶⁹ Kirk & Raven, p. 281.

⁷⁰ K. Reich, "Parmenides und die Pythagoreer," *Hermes*, vol. 82 (1954), pp. 287–294; N. B. Booth in *Phronesis*, vol. 2 (1957), pp. 92–99. On the other hand a reference to Heraclitus in fr. 6.9 is fairly certain, see W. Kranz, *Rheinisches Museum*, vol. 101 (1958), pp. 250–254; Vlastos (as against Kirk) in *American Journal of Philology*, vol. 76 (1955), p. 341, n. 11 and pp. 348–353; and Guthrie, *History*, vol. 1, p. 468.

⁷¹ "Parmenides on Thought and Being," *Philosophical Review*, vol. 64 (1955), pp. 546–560.

⁷² I do not deal with discussions of the Proem—see E. A. Havelock in *Harvard Studies in Classical Philology*, vol. 63 (1958), pp. 133–143, and K. Deichgraber, "Parmenides' Auffahrt zur Göttin des Rechts," *Abhandlungen Akademie der Wissenschaften zu Mainz* (Wiesbaden, 1958). For Parmenides and Logic see G. Jameson, "Well-rounded Truth and Circular Thought in Parmenides," *Phronesis*, vol. 3 (1958), pp. 15–30, and J. Stannard in *Philosophical Review*, vol. 69 (1960), pp. 526–533. For existentialist interpretations see Beaufret, n. 23 above, and for a Hegelian view W. Broecker, "Gorgias contra Parmenides," *Hermes*, vol. 86 (1958), pp. 429–430.

⁷³ N. B. Booth, "Were Zeno's Arguments Directed Against the Pythagoreans?" *Phronesis*, vol. 2 (1957), pp. 90–103; "Were Zeno's Arguments a Reply to Attacks upon Parmenides?" *Phronesis*, vol. 2 (1957), pp. 1–9; G. E. L. Owen, "Zeno and the Mathematicians," *Proceedings of the Aristotelian Society*, vol. 58 (1957–58), pp. 199–222; W. Kullmann, "Zenon und die Lehre des Parmenides," *Hermes*, vol. 86 (1958), pp. 157–172; Vlastos, *Philosophical Review*, vol. 68 (1959), pp. 532–535. Raven maintains that the arguments are directed particularly against the Pythagoreans—Kirk & Raven, pp. 290–291.

⁷⁴ G. Ryle, *Dilemmas* (Cambridge, 1954), chap. iii, "Achilles and the Tortoise"; J. F. Thomson in *Analysis*, vol. 15 (1954), pp. 1–13; D. S. Schwyder in *Journal of Philosophy*, vol. 52 (1955), pp. 449–459; V. C. Chappell, *ibid.*, vol. 59 (1962), pp. 197–213. See also Max Black, *Problems of Analysis* (Ithaca, 1954), p. 109, n. 1.

⁷⁵ See n. 24 above.

⁷⁶ See N. B. Booth, "Zeno's Paradoxes," *Journal of Hellenic Studies*, vol. 77 (1957), pp. 187–201; Vlastos in *Gnomon*, vol. 31 (1959), pp. 195–199 and the discussion by him in W. Kaufmann, *Philosophic Classics* (New York, 1961), pp. 27–45, with further references on p. 28 n. 4.

9. MELISSUS. Interest has always centered around two supposed developments in Eleatic doctrine attributed to Melissus: (1) His view that the One Being is infinite in extent as well as in time. This doctrine has now been denied by Vlastos,⁷⁷ but without good grounds, it would seem.⁷⁸ (2) The view that the One Being is incorporeal. For this the evidence seems unimpeachable.⁷⁹ But it has been attacked once again,⁸⁰ although I would judge without success. There have been interesting contributions on points of detail as well,⁸¹ but there is some truth in the contention that on the whole Melissus has not received the attention he deserves which is made by J. H. M. Loenen,⁸² although probably few will follow him in the radically new interpretations which he offers.

10. EMPEDOCLES. The considerable literature devoted to discussion of details⁸³ can hardly be said to have contributed much to the solution of major problems, with one exception. On the compatibility of the two poems Raven, after a careful discussion, concludes⁸⁴ that Empedocles regarded the migratory soul and physical consciousness as quite distinct in themselves, and yet the former had some sort of physical basis as being an attunement or harmony involving the four elements together with Love and Strife. The problem of the two poems has been further discussed in an interesting and valuable article by Kahn.⁸⁵ Zafiropulo has sought to incorporate Empedocles within his general view of the development of the Presocratics,⁸⁶ and to this may be added the general survey by G. Nélod,⁸⁷ which is perhaps intended more for the general reader than for the specialist.

11. ANAXAGORAS. It has often been held that

the system of Anaxagoras is of all the Presocratic world-pictures the most difficult to reconstruct satisfactorily. A succession of major but divergent modern reconstructions might seem to bear this out. Anaxagoras seems clearly to have two positions or contentions attributed to him: (1) that everything contains a portion of everything else (Universal Mixture), and (2) that things consist of parts that are like the whole and like one another (Homoeomereity). The difficulty has been to relate these two contentions to each other. It may be held (a) that they were in fact inconsistent in Anaxagoras' thought, and he either noticed this or failed to notice it; (b) that they are somehow fully consistent; (c) that Universal Mixture was only true in some limited or qualified sense, e.g., only seeds or opposites have all qualities and things are constituted in the first instance out of these or grow out of them; (d) that Homoeomereity must be limited or sacrificed in some way in order to preserve Universal Mixture. For a long period variants under (c) have seemed to scholars likely to provide a solution, and here belong the reconstructions by Raven,⁸⁸ and C. Mugler.⁸⁹ However, all such solutions involve a two-stage analysis of matter. There is no reference to this in any ancient author; Aristotle is emphatic, it would seem clear in his own mind that for Anaxagoras in each *thing* there were portions of every substance; and finally Anaxagoras' own formula for Universal Mixture has no qualification of this kind attached, although it has a qualification about *Nous*. This suggests that it had no other qualification apart from the case of *Nous*. Views involving (d) are presented by R. Matthewson,⁹⁰ and M. E. Reesor.⁹¹ In all such reconstructions it is probable that the reader will give assent more readily to the criticisms

⁷⁷ *Gnomon*, vol. 25 (1953), pp. 34-35.

⁷⁸ So Raven in Kirk & Raven, p. 300.

⁷⁹ Raven, *ibid.*, pp. 302-304; Vlastos, n. 78 above.

⁸⁰ N. B. Booth, "Did Melissus Believe in Incorporeal Being?" *American Journal of Philology*, vol. 79 (1958), pp. 61-65.

⁸¹ So Kirk and Stokes, *Phronesis*, vol. 5 (1960), pp. 1-4; G. E. Gershenson and D. A. Greenberg, *ibid.*, vol. 6 (1961), pp. 1-9.

⁸² See p. 124 of the book mentioned in n. 63 above.

⁸³ Special mention may be made of N. B. Booth, "Empedocles' Account of Breathing," *Journal of Hellenic Studies*, vol. 80 (1960), pp. 10-15 and H. A. T. Reiche, *Empedocles' Mixture, Eudoxan Astronomy, and Aristotle's Connate Pneuma* (Amsterdam, 1960).

⁸⁴ Kirk & Raven, pp. 335-360.

⁸⁵ "Religion and Natural Philosophy in Empedocles' Doctrine of the Soul," *Archiv für Geschichte der Philosophie*, vol. 42 (1960), pp. 3-35. See also H. Munding, "Zur Beweisführung des Empedokles," *Hermes*, vol. 82 (1954), pp. 129-145.

⁸⁶ N. 41 above.

⁸⁷ *Empédocle d'Agrigente* (Brussels, 1959).

⁸⁸ *Classical Quarterly*, vol. 48 (1954), pp. 123-137, and in Kirk & Raven, pp. 367-394.

⁸⁹ *Revue des Études Grecques*, vol. 49 (1956), pp. 314-376.

⁹⁰ "Aristotle and Anaxagoras," *Classical Quarterly*, vol. 52 (1958), pp. 67-81.

⁹¹ "The Meaning of Anaxagoras," *Classical Philology*, vol. 55 (1960), pp. 1-8.

made of other reconstructions than to the positive proposals. No final solution is in sight, but a clue to the ultimate answer may lie in Vlastos' contention,⁹² despite its subsequent rejection by C. Strang,⁹³ that there need be nothing vicious in a sequence such that gold, while containing a portion of everything else, has a predominant ingredient (gold) which, while containing a portion of everything else, has a predominant ingredient (gold) which. . . . The objection usually made to this is that it gives no starting point for building larger units out of primary constituents. But if the question posed was how to analyze phenomenal objects already in existence this objection might not apply. This would open the way for an interpretation under (b) above. Apart from this fundamental question there have been important discussions of points of detail,⁹⁴ and the method of formulation of the world has been studied by D. Bargrave-Weaver.⁹⁵

12. ATOMISM. DEMOCRITUS. The system of Democritus has been studied by V. E. Alfieri in a book with a somewhat misleading title, *Atomos Idea, l'Origine del concetto dell' atomo nel pensiero greco*,⁹⁶ which is not concerned with the origins of atomism in the historical sense.⁹⁷ Accepting that the logical starting point for atomism was the Eleatic position, he deals with most of the problems concerning Democritus, and has important though controversial things to say about the doctrine of the soul. In general fundamental studies of atomism in this period are fewer than one might have hoped for

and than the importance of the school calls for. The treatment in Kirk and Raven is extremely brief, and no one as far as I am aware has discussed fully the problem of perception in the atomic system. The one really important contribution to the study of atomism is probably that of J. Mau,⁹⁸ discussing the question of how and when the problem of infinitesimals was first raised—whether before Zeno or not, and whether it was in connection with the Pythagorean problem of incommensurability. He adopts an intermediate position between the two main extreme views and then carries the story of infinitesimals down to Epicurus. C. Mugler has written on theories of life and conscience in Democritus,⁹⁹ and D. McGibbon on Pleasure.¹⁰⁰

13. THE SOPHISTS. With these I can deal only briefly. Two parts of Untersteiner's commentary on the texts fall in this period.¹⁰¹ On Protagoras, Kerferd has discussed the consistency of the positions apparently ascribed to him by Plato,¹⁰² and R. F. Holland has protested against the manner of approach adopted by Untersteiner to the problem of the Man-measure doctrine.¹⁰³ The whole range of evidence for Protagoras is sensibly dealt with by A. Capizzi,¹⁰⁴ and by S. Zeppi,¹⁰⁵ and there is a valuable summary treatment by K. von Fritz.¹⁰⁶ For discussion of the evidence of the *Theaetetus* one may mention the articles by A. Capelle,¹⁰⁷ and L. Versenyi.¹⁰⁸ More general in approach is G. M. Sciacca's book, *Gli Dei in Protagora*, which does not deal only with Prota-

⁹² *Philosophical Review*, vol. 59 (1950), p. 51.

⁹³ "The Physical Theory of Anaxagoras," *Archiv für Geschichte der Philosophie*, vol. 45 (1963), pp. 101–102. Strang's own interpretation falls under (c) in that he maintains that elements do not contain a portion of everything else, although phenomenal objects do. But his elements can never be isolated, and so seem to have logical rather than physical existence. Thus (d) also is in effect preserved. The trouble is, as so often, that there is nothing about this in ancient sources.

⁹⁴ E.g., A. Wasserstein, "A note on fr. 12," *Classical Review*, vol. 10 (1960), pp. 4–5.

⁹⁵ "The Cosmogony of Anaxagoras," *Phronesis*, vol. 4 (1959), pp. 77–91.

⁹⁶ Florence, 1953.

⁹⁷ For this see O. Luschkat, "Wie das Atom erdacht werde, Versuch einer Entstehungsgeschichte des antiken Atomismus," *Forschungen und Fortschritte*, vol. 27 (1953), pp. 136–141, and W. Kranz, "Die Entstehung des Atomismus," *Convivium (Festschrift Ziegler)* (Stuttgart, 1954), pp. 14–40.

⁹⁸ *Zum Problem des Infinitesimalen bei den antiken Atomisten* (Berlin, 1954).

⁹⁹ *Revue de Philologie*, vol. 33 (1959), pp. 7–38.

¹⁰⁰ *Phronesis*, vol. 5 (1960), pp. 75–77.

¹⁰¹ See n. 24 above.

¹⁰² "Protagoras' Doctrine of Justice and Virtue in the 'Protagoras' of Plato," *Journal of Hellenic Studies*, vol. 73 (1953), pp. 42–45. See also G. Vlastos' *Introduction* to the Liberal Arts Press Translation of *Plato's Protagoras* (New York, 1956).

¹⁰³ *Classical Quarterly*, vol. 7 (1956), pp. 215–220.

¹⁰⁴ *Protagora, le testimonianze e frammenti* (Florence, 1955).

¹⁰⁵ *Protagora e la filosofia del suo tempo* (Florence, 1961).

¹⁰⁶ S.v. *Protagoras*, Pauly-Wissowa, *Realencyklopädie*, vol. 45 (1957), pp. 908–921.

¹⁰⁷ *Hermes*, vol. 88 (1960), pp. 265–280.

¹⁰⁸ *American Journal of Philology*, vol. 83 (1962), pp. 178–184.

goras.¹⁰⁹ On Prodicus two discussions dealing with his ethical doctrines may be mentioned,¹¹⁰ and one summary article.¹¹¹ In the case of Gorgias, I omit works dealing primarily with literary and rhetorical aspects and doctrines.¹¹² Treatments of the arguments in his work *On that which is not* have been provided by V. di Benedetto,¹¹³ Kerferd,¹¹⁴ Loenen,¹¹⁵ and W. Broecker.¹¹⁶ G. Calogero has argued on rather weak evidence for a link with Socrates.¹¹⁷ The edition of the fragments by C. Moreschini¹¹⁸ is rather slight. In the case of Antiphon controversy has continued as to whether there were two Antiphons or only one,¹¹⁹ and

interpretations of his doctrines have been offered by Kerferd,¹²⁰ and J. S. Morrison.¹²¹ For other sophists I mention Max Salomon¹²² and G. F. Hourani¹²³ on Thrasymachus, and B. Wisniewski¹²⁴ and E. S. Ramage¹²⁵ on the *Dissoi Logoi* as raising points of interest. In general it may be said that while, *pace* Kirk and Raven,¹²⁶ there is a growing recognition of the philosophic importance of the major sophists, this has yet to be reflected in any large scale treatment which will present the detailed evidence, and not at the same time give a distorted or eccentric view of the movement as a whole.

University College, Swansea, Wales

¹⁰⁹ Palermo, 1958.

¹¹⁰ G. B. Kerferd, "The Relativism of Prodicus," *Bulletin of the John Rylands Library*, vol. 37 (1954-55), pp. 249-256, and S. Zeppi, "L'etica di Prodicco," *Rivista Critica di Storia di Filosofia*, vol. 11 (1956), pp. 265-272, reprinted in his *Studi sulla Filosofia Presocratica* (Florence, 1962).

¹¹¹ K. von Fritz, s.v. *Prodikos*, Pauly-Wissowa, *Realencyklopädie*, vol. 45 (1957), pp. 85-89.

¹¹² But Dodds' edition of *Plato's Gorgias* (Oxford, 1959) must be mentioned.

¹¹³ *Rendiconti della Reale Accademia dei Lincei*, vol. 10 (1955), pp. 287-307.

¹¹⁴ *Phronesis*, vol. 1 (1955), pp. 3-25.

¹¹⁵ N. 63 above.

¹¹⁶ *Hermes*, vol. 86 (1958), pp. 425-440.

¹¹⁷ "Gorgias and the Socratic Principle *nemo sua sponte peccat*," *Journal of Hellenic Studies*, vol. 77 (1957), pp. 12-17.

¹¹⁸ Turin, 1959.

¹¹⁹ J. S. Morrison, *Classical Review*, vol. 3 (1953), pp. 3-6; E. R. Dodds, *ibid.*, vol. 4 (1954), pp. 94-95; J. S. Morrison, *Proceedings of the Cambridge Philology Society*, vol. 7 (1961), pp. 49-58.

¹²⁰ *Proceedings of the Cambridge Philology Society*, vol. 4 (1956-57), pp. 26-32.

¹²¹ *Phronesis*, vol. 8 (1963), pp. 35-49.

¹²² *Zeitschrift für Philosophische Forschung*, vol. 7 (1953), pp. 481-492.

¹²³ *Phronesis*, vol. 7 (1962), pp. 110-120.

¹²⁴ *L'Antiquité Classique*, vol. 28 (1959), pp. 80-97, and *Classica et Mediaevalia*, vol. 12 (1961), pp. 106-116, both articles being nominally concerned with Hippias.

¹²⁵ *American Journal of Philology*, vol. 82 (1961), pp. 418-424.

¹²⁶ N. 22 above.

V. BASIC ACTIONS

ARTHUR C. DANTO

"Well, why should we want to know?" said Verity, giving a yawn or causing herself to give one.

I. Compton-Burnett, *Two Worlds and Their Ways*

I

"THE man *M* causes the stone *S* to move." This is a very general description of a very familiar sort of episode. It is so general, indeed, that it does not tell us whether or not *M* has performed an action. The description holds in either case; so it *could* have been an action. Without pausing to inquire what further features are required for it definitely to have been an action, let us merely note that *there are* actions that fall under the general description of "causing something to happen." Yet, since this description leaves it unclear whether or not an action has been performed, performing an action cannot be one of the truth conditions for "causing something to happen." And since this description cuts across those two cases, we may assume we are employing the same sense of the expression "causes something to happen" in both. Presumably, we are using "causes" in just the same sense whether we say that the man *M* causes the stone *S* to move or we say that the stone *S* causes the pebble *P* to move. If it is clear from the latter sentence that an action has *not* been performed, this clarity will be due to certain facts about stones rather than to any difference in the concept of causality. It is commonly assumed that stones never perform actions, although men sometimes do. Hence the indefiniteness of our original sentence is not due to any ambiguity in the concept of causality, but rather to certain facts about men, or to certain assumed facts. The concept of causality allows us to ignore differences between men and stones, as well as differences between performing an action and not.

I shall persist in speaking of *individuals* (the man *M*, the stone *S*) causing things to happen, even though our concept of causality has been classically analyzed as a relationship between pairs of *events*. According to the classical analysis, the movement of the pebble *P* is one event, the effect of another event, which I shall, with studied ambiguity,

simply designate an *S*-event, in this case its cause. Comparably, the movement of *S* in my other example is one event, the effect of another event, similarly and no less ambiguously to be designated an *M*-event, which is its cause. And this *M*-event, whether or not it is an action performed by *M*, is correctly (if rather generally) to be described as *causing something to happen*—namely, the movement of *S*.

I shall now suppose that my original sentence in fact describes an action performed by *M* (moving the stone *S*). Of this particular spatial translation of *S* we may say three distinct and relevant things: that it is (*a*) an action, performed by *M*; that it is (*b*) something that was *caused* to happen (in this case by *M*); and that it is (*c*) the effect of an event distinct from itself (in this case the *M*-event). That this event can be both (*a*) and (*b*) follows from the remarks in the first paragraph. That—disregarding the special information in parentheses—(*c*) must hold if (*b*) does—follows from the analysis of causality referred to in the second paragraph. That it is (*b*) follows, I suppose, from the fact that *S* is a stone: stones don't *just* start to move without something causing them to move.

We must now look into the *M*-event itself. Do all three characterizations apply to *it*? This, I fear, cannot be decided without investigation. Let us suppose, however, that the *M*-event is both (*a*) and (*b*), for it might well be. Then it must also be (*c*), and there must then be yet another event, distinct from it, which is its cause. This may be yet a further *M*-event, and about it we may raise the same question. It would be rash to claim that we have slid into an infinite regress, damaging or otherwise. But if a given *M*-event is both (*a*) and (*b*) and, hence, (*c*), then ultimately its being (*c*) must lead us to a further *M*-event, which is (*a*) and *not* (*b*). And unless some *M*-events are (*a*) and *not* (*b*), *no M*-events are ever (*a*). That is, if there are any actions at all, there must be two distinct *kinds* of actions: those performed by an individual *M*, which he may be said to have *caused* to happen; and those actions, also performed by *M*, which he cannot be said to have caused to

happen. The latter I shall designate as *basic actions*.

In this paper, I shall defend (and explore the consequences of) four theses which I regard as fundamental to the theory of action:

- (1) If there are any actions at all, there are basic actions.
- (2) There are basic actions.
- (3) Not every action is a basic action.¹
- (4) If *a* is an action performed by *M*, then either *a* is a basic action of *M*, or else it is the effect of a chain of causes the originating member of which is a basic action of *M*.

I wish first to make quite clear the sense in which an individual does not cause his basic actions to happen. When an individual *M* performs a basic action *a*, there is no event distinct from *a* that both stands to *a* as cause to effect and is an action performed by *M*. So when *M* performs a basic action, he does nothing first that causes it to happen. It will be convenient to consider two possible objections to this.

It may be objected, first, that there are or may be other senses of "causes" than the sense mentioned above, in accordance with which it would be proper to say that *M* causes his basic actions to happen. Thus, if raising an arm were an instance of a basic action, an individual who does this might still be said to cause it to happen in some sense of "cause" other than the sense that I reject in application to basic actions. I accept this objection: there *may be* such other senses of "cause." But (i) we should still require exactly the same distinction that I am urging within the class of actions, and I should therefore be defending the *verbally* distinct thesis that unless there were actions an individual causes to happen in this *new* sense, there would be no actions he caused to happen in the original sense, either. So, unless there were actions of the former sort, causing a stone to move would, for example, never be an *action* that anyone performed (although men might still cause stones to move, since performing an action is not a truth-condition for "causing something to happen"). And (ii) this new sense of "cause" would *not* apply *whether or not* an action had been performed. It should, indeed, be absolutely clear from the sentence "*M* caused *a* to happen"—using this special sense of "cause"—that *M* had performed an action. Those who find it convenient to maintain that the concept of causality is invariant to the distinction between performing an action and not, would have as little use for this new sense of "cause" as I do. Neither

they nor I would want to say that *stones* cause *anything* to happen in this new sense of "cause." Not that I wish to restrict the performance of basic actions to men alone. Other individuals may, for all I know, perform them as well. Some theologians have spoken as though everything done by God were a basic action. This would prohibit us, of course, from saying that God caused anything to happen (the making of the Universe would be a basic action.) And, for reasons which will soon emerge, this would make the ways of God inscrutable indeed.

It may be objected, second, that if we take the absence of a cause to be the distinguishing mark of a basic action, then we must class as basic actions a great many events that we should be disinclined, on other grounds, to accept as actions at all, e.g., the uniform rectilinear motion of an isolated particle, or perhaps any instance of radioactive decay. This objection is readily deflected. I have not claimed that basic actions are not caused, but only that a man performing one does not cause it by performing some other action that stands to it as cause to effect. Moreover, the absence of a cause would not be a sufficient criterion for a basic action, even if basic actions *were* uncaused. It would serve only to mark off a special class of actions from the rest. Of course, only what is already an action can be a *basic* action. And I have not so much as tried to say what are the general criteria for actions.

II

I have avoided citing unconditional instances of basic actions, in part because any expression I might use, e.g., "moving a limb," could also be used to designate something that was caused to happen, or something that was not an action, much less a basic one. I think there is nothing that is always and in each of its instances an unmistakably basic action. This is reflected by language in the fact that from the bare description "*M*'s limb moved," for example, one could not tell whether *M* had performed a basic action or even an action. Nor could one tell this by observing only the motion of the limb without bringing in differentiating contextual features. I have accordingly contented myself with the neutral expression "*M*-event," declaring it to be a basic action when I required an instance.

Now I wish to specify some of the differentiating contextual features, and I shall consider four

¹ Thesis (3) is explored in detail in my paper, "What We Can Do," *The Journal of Philosophy*, vol. 60 (July, 1963), pp. 435-445.

distinct cases, all of which might indifferently be covered by the same description, so that the description alone leaves it unclear whether an action has been performed or not. Of the four cases, three (*C-1*, *C-2*, *C-4*) will indeed be actions, and of these one (*C-4*) will be a basic action. The four cases together might be termed a *declension* of the description. Not every such description admits of the full declension, for some appear never to be exemplified as basic actions at all. "Moving a stone," I should think, never, or not ordinarily, is exemplified as a basic action, though we have seen that it may be exemplified by an action. I want to begin with a deliberately controversial example and shall decline the expression "*M* laughs."

C-1. M causes himself to laugh. I am thinking here of cases where someone does something to make himself laugh, and does not simply laugh because of something he happens to do. Thus I may do something ridiculous and laugh because I find it so, but I did not do this ridiculous thing in order to make myself laugh. Again, I sniff a cartridge of nitrous oxide, not knowing it to be nitrous oxide, but just to find out what it is. But, since it is nitrous oxide, I laugh, though I did not sniff to make myself laugh. I wish to include only cases where I do something ridiculous or sniff from a private cartridge of nitrous oxide *in order to* laugh, perhaps because I think laughter good for the liver or because I just enjoy laughing and cannot always wait for someone or something to come along and cause me to laugh. I definitely want to exclude a comedian who laughs at some reruns of his antic films (unless he had them rerun for this special purpose), and definitely want to include someone who deliberately engages in auto-titillation to excite spasmodic laughter. Doubtless, episodes falling under *C-1* are rare in normal adults in our culture, but this is irrelevant. Also irrelevant is the fact that people don't laugh *at* the nitrous oxide they sniff, though they do laugh at the silly faces they pull, for their own delectation, in mirrors.

C-2. Someone or something other than M causes M to laugh. This is the typical case for adults and children in our culture. It is for my purposes again irrelevant whether the cause of *M*'s laughter is also its object, or whether it has an object at all (as it does not if he is tickled or submitted to nitrous oxide). Similarly, it is irrelevant whether, in case someone causes *M* to laugh, the former has performed an action or not, whether, that is, he did what he did in order to make *M* laugh. For it is what *M* does that uniquely concerns us here.

C-3. M suffers a nervous disorder symptomized by spasmodic laughter. This is comparable, say, to a tic: *M* laughs unpredictably, and for "no reason." Such laughter is mirthless, of course, but so are some instances falling under the two first cases. It may be argued that the entire case falls under *C-2*, and that in identifying it as the symptom of a nervous disorder, I have marked off a class of causes for *M*'s laughter. Still, the case requires special consideration, in that *M*'s laughing here is never an action, whereas his laughter under *C-2* sometimes is.

C-4. M has the true power of laughing. By this I mean that *M* laughs when he wants to without (in contrast with *C-1*) having to cause himself to laugh; without (in contrast with *C-2*) someone or something having to cause him to laugh; without, finally, as in *C-3*, suffering from the relevant nervous disorder. This does not mean that *M* is normal, but only that his abnormality is of a benign sort; i.e., it is by way of a gift. His laughing may have an object: he may, when he wishes, direct a stream of laughter at whom or what he chooses, without the chosen object ever being a cause of his laughing.

Instances falling under *C-4* are perhaps rare, but these alone would qualify as basic actions performed by *M* when "*M* laughs" is true. I have identified the case not so much by specifying what differentiating contextual features must be present, but by specifying what differentiating contextual features must be *absent*. Notice that *M*'s laughing here differs markedly from the ability most of us have of making laugh-like noises, e.g., for the sake of politeness, or to save our reputation for seeing a joke when we don't see it, or to play a mocker's role in an amateur theatrical. Most of us can pretend so to laugh: but I speak here of laughing, not of "laughing."

I want now to comment on these four cases.

When *M* laughs under *C-1*, we may say of his laughing three distinct things: that it is (a) an action of *M*'s; that it is (b) something that *M* causes to happen; and that it is (c) the effect of some event, distinct from itself (an *M*-event) which is its cause. *M*'s laughing here is an action in just the same sense in which his causing a stone to move is an action. Causing himself to laugh is the action he performed, though of course the description "*M* caused himself to laugh" leaves it unclear, as in the case of the stone, whether he performed an action at all. One could mark that difference only by bringing in the general differentiating features of action.

In C-2, *M* does not cause himself to laugh, and one may find reasons for balking at the claim that his laughing, in such a case, is an action of his at all. For consider this argument. When *M* causes a stone *S* to move, we may agree that the action is *M*'s. But we reject the claim that it is an action of *S*'s. So parity suggests that when someone moves *M* to laughter, this may be an action performed by the former, but not an action of *M*'s.

What I must do is to show that parity is inoperative, and so justify my claim that instances of C-2 are actions in contrast with instances of C-3. Well, I shall somewhat artificially suggest that *M*'s action here requires this description: what he does is to *not not* laugh. The double negative is not, in the language of action, a triviality. Logically, of course, the double negative of a proposition is just that proposition, and from a strictly logical point of view, we could say the same thing, albeit more awkwardly, with "The man *M* causes the stone *S* to not not move" as we straightforwardly say with "The man *M* caused the stone *S* to move." I wish, in fact, to retain that regular inferential feature of double negation which allows us to proceed from not not *A* to *A*, but for the case of action I wish to exclude the reverse inference. For my double negative marks the case of *negligence*, and whether a negligence is to be ascribed to someone is a case for independent investigation. So, pending such investigation, we cannot say, on the basis of knowing that a man laughs, that he is to be charged with negligence. And for this reason we cannot automatically go from "laughs" to "not not laughs." Indeed, since we don't ascribe negligence to stones, it would be invalid, given my convention, to proceed from "the stone moves" to "the stone not not moves."

Do we quite want to say, then, that C-2 is to be restated thus: *Someone or something other than M causes M to not not laugh*? Perhaps we would, in spite of flaunting usage. What we would be saying, however, is only this: that *M* was excited to laugh and did nothing to inhibit his laughter. And it is our common assumption that men are normally capable of doing something which, in effect, stops the flow of laughter from issuing forth in, say, public guffaws. Whether men are called upon to exercise these inhibitory practices varies from context to context: in the music hall there is license to suspend them, to "let oneself go," but at High Mass there is not. It is in such contexts only that laughter is *pronounced* a negligence, but blaming, surely, does not make of something an action

when it would not otherwise have been so. It is only insofar as something is an action already that blaming it, or blaming someone for doing it, is appropriate.

With regard to C-3, however, the laughter stands liable to no special charge of negligence: his laughing fails to be a case of not not laughing, for identification of it as a nervous disorder, or in the syndrome of one, locates it beyond the control of the man who is so afflicted. It is, indeed, almost a paradigm case of this: like a hiccup. One *might* blame the man for being in a place where his symptom, easily mistakable as a negligence, might break out unpredictably. Or we might blame him again for a kind of negligence in "not doing something about it," viz., going to a nerve specialist, assuming there is a known cure. At all events, it is plain enough why C-3 differs from C-2. The critical issue, of course, is the matter of *control*, and this brings us to C-4. And the rest of this paper is by way of a comment on C-4.

Most readers, I think, will resist the suggestion that C-4 is a case of action. There is good reason for this. For most of us, laughing as a *basic action* is unintelligible. I shall hope to show why this is so, and showing it will involve a demonstration of thesis (2). Meanwhile, the reader might ponder the precise analogue to this in the case of *moving an arm*, which admits of a full declension. Thus C-1: *M* causes his arm to move, i.e., by striking it with his other arm; C-2: someone or something other than *M* causes *M*'s arm to move, e.g., by striking it; C-3: *M* suffers from a nervous disorder, so his arm moves spasmodically and unpredictably, as a kind of tic; and C-4: *M* moves his arm without suffering from a nervous disorder, without someone or something causing it to move, without having to do anything to cause it to move. Here, I am certain, C-4 is the *typical* case. Moving an arm is one of the standard basic actions. If we now seek to determine in what way this behavior is intelligible, we should have no great difficulty in seeing why laughing under C-4 is *not*.

III

Suppose now that moving a stone is an action performed by *M*. It is difficult to suppose that *moving a stone* admits of a full declension, largely because it seems to lack cases for C-3 and C-4. In fact there are difficulties in finding instances for C-1 and C-2 unless we change the sense of possession (*M*'s arm, *M*'s stone) from philosophical to legal ownership. But for the moment I shall be

concerned only with the fact that we move stones only by causing them to move. This then means that, in order to cause the motion of the stone, something else must be done, or must happen, which is an event distinct from the motion of the stone, and which stands to it as cause to effect. Now this other event may or may not be a basic action of *M*'s. But if it is not, and if it remains nevertheless true that moving the stone is an action of his, then there must be something else that *M* does, which causes something to happen which in turn causes the motion of the stone. And *this* may be a basic action or it may not. But now this goes on forever unless, at some point, a basic action is performed by *M*. For suppose every action were a case of the agent causing something to happen. This means, each time he does *a*, he must independently do *b*, which causes *a* to happen. But then, in order to do *b*, he must first independently do *c*, which causes *b* to happen. . . . This quickly entails that the agent could perform no action at all. If, accordingly, there are any actions at all of the sort described by "causing something to happen," there must be actions which are *not* caused to happen by the man who performs them. And these are basic actions.

But this argument is perfectly general. If there are any actions at all, there are basic actions. This is a proof of thesis (1). Moreover, if *M* performs an action describable by "causing something to happen," he must also, as part of what he does, perform an action that he does not cause to happen. And this is a proof of thesis (4). It would be a proof of thesis (2) if in fact there were actions described as "causing something to happen." This would then require us to accept thesis (3) as true: for such an action would not be a basic action, and so not every action is basic.

I do not wish to suggest, however, that the only proof we are entitled to, for the existence of basic actions, is by way of a transcendental deduction, for I believe we all know, in a direct and intuitive way, that there are basic actions, and which actions are basic ones. To show that we do know this will clarify one of the ways in which laughing is a controversial instance of a basic action.

I must make a few preliminary remarks. First, every *normal person* has just the same *repertoire R* of basic actions, and having *R* is what defines a normal person for the theory of action. Second, persons may be *positively abnormal* when their repertoire of basic actions includes actions not included in *R*, and may be *negatively abnormal* when actions included in *R* are not included in their repertoire. Some persons may be both positively and negatively abnormal, e.g., someone who laughs as a basic action but who is paralyzed in one arm. If someone's repertoire is empty, he is capable of no basic actions, and hence of no actions. Such a deprived entity is a *pure patient*, e.g., like a stone. Plainly, our repertoire of actions is greater than our repertoire of basic actions, though a being who performed every possible action and all of whose actions were basic actions may be conceived of: such a being would be a *pure agent*. For the present, however, I am concerned with beings intermediate between pure patients and pure agents, and I want now to say that basic actions are *given* to such beings in two distinct senses, each of which bears a definite analogy to a sense that the term has in the theory of knowledge.²

(i) In the theory of knowledge, to say that *p* is *given* is in part to point a contrast: one is saying that *p* is not inferred from some other proposition. Analogously, when I speak of an action as given, I shall mean to say, in effect, that it is a basic action, and point a contrast with actions we *cause* to happen. The notion of givenness is understood this way: *p* is a starting point for an inference to another and (commonly) different proposition *q* for which *p* provides at least part of the evidence. Analogously, an action *a*, if a basic action, is a starting point for the performance of another action *b*, of which it is at least part of the cause. "Is caused by" and "is inferred from" are analogous relations in the theories of knowledge and of action, respectively.

(ii) It has been argued that the distinction between *basic sentences* and sentences of other kinds is not ultimate, that a sentence which, in one context, is indeed a starting point for an inference to another, may, in a different context, itself be inferred to, and hence an end point in an in-

² The analogy between theory of knowledge and theory of action runs very deep indeed, almost as though they were isomorphic models for some calculus. Obviously, there are things we can say about actions that do not hold for cognitions, etc., but this means very little. Suppose we have two models *M-i* and *M-j* for a calculus *C*, and suppose that "star" plays in the same role in *M-i* that "book" plays in *M-j*. It is hardly an argument against their both being models for *C* that we don't print stars or that books are not centers of solar systems. I shall use theory-of-knowledge features as a guide for structuring the theory of action. When the analogy gives way, it will be interesting to see why it does.

ference.³ Analogously, an action *a* may, in one context, be a starting point and basic, while it may be caused to happen in a different one. There is some justice in this latter claim: as we have seen, one cannot tell from the bare description "moving an arm" whether a basic action is referred to, or even an action. But, thinking now of sentences, perhaps some restriction can be put on the *kind* of sentence which can be given in sense (i). If *p* is given in one context and inferred in another, there might nevertheless be sentences which are never basic and always are inferred. And a corresponding restriction might hold in the theory of action: even if any action that is ever basic might, under a sufficiently general description, be caused to happen in another context, there might be actions that never are basic under any description. In the theory of knowledge, one such restriction is often defended, namely that basic sentences are those and only those which can be conclusively verified by sense experience, and that no other kind of sentence ever can be given. But within the class of potentially given sentences, a division might be made along the customary lines of sense-modality, i.e., those verified by seeing, or by audition, or by touch, etc. We might then define an *epistemically* normal person as one who experiences in all modes. A negatively abnormal person would then be deficient in at least one such mode, e.g., is blind; and a positively abnormal person then experiences in some mode outside the normal repertoire, e.g., has some "sixth sense." The analogy to the theory of action is obvious. But by means of it we may introduce our second sense of given: the normal modes of experience are "given" in the sense that they constitute the standard cognitive equipment. The normal person has various classes of starting points for inferences as he has various classes of starting points for actions. These are given in the sense that they are not for the most part *acquired*. Thus we speak of the "gift of sight," etc. This does not mean that there need be any sentences in the superstructure to which a negatively abnormal person might not infer: he is deficient only at the base: and then not *totally* deficient (or if he is, then he cannot have any empirical knowledge, is *cognitively impotent*). And similarly, *toutes proportions gardées*,

with the negatively abnormal person as defined in the theory of action.

Now when a blind man says that he can know whether a certain object is red or not, there are two senses or uses of "can" that are compatible with his abnormality. He must mean either that he can *infer* to "*x* is red" from other sentences or that his case is not medically hopeless, that by means of a cure he may be restored to that state of normality in which such sentences may be known by him directly and not, as it were, *merely* by means of inference. Yet there is a true and in fact an *analytic* sense in which a blind man cannot know whether a certain object is red, nor, on certain accounts of meaning, so much as know what such a sentence *means* (the non-analytic senses are usually false). The situation of a *paralyzed* man is perfectly analogous. When he sincerely says that he can move his arm, he must mean either that he can *cause* it to move, or that his situation is not medically hopeless. But, in again a true and an analytical sense, he cannot move his arm and does not know, does not so much as understand, what it means to move his arm in the way in which a normal person understands this. For this is the kind of understanding that is alone given to those who have the power to move their arms in the normal, basic way. This kind of understanding cannot so much as be conveyed to a negatively abnormal person while he is so.

Some of the chief difficulties philosophers have encountered in the theory of action are due to their having approached it from the point of view of the negatively abnormal. From *that* point of view, basic action is hopelessly mysterious. There is, however, perhaps no better way of eliciting the quality of our knowledge of these things than to think of endeavoring to remove the mysteriousness surrounding these actions in the thwarted comprehension of the negatively abnormal person. We may achieve some sympathy for his plight by imagining *ourselves* similarly confronting someone who is *positively* abnormal, who can perform, as a basic action, what we at best can cause to happen, and then asking *him* to give us an understanding of his gift. The fact is that we cannot explain to the negatively abnormal, nor can the positively

³ Though not always without some awkwardness. Suppose it were held that only sentences can be given which have the form of first-person reports of sense-experience, e.g., "I now see a reddish *x* . . ." Such a sentence is not easily rendered as the conclusion of an inference, though it can be so rendered, I suppose, if I both knew that something *x* had an unmistakable taste and that whatever has this taste is red. Then, by tasting *x* and seeing only its silhouette, I might feel secure in inferring that I was seeing a reddish *x*. Of course there are philosophically crucial senses of "see" which would rule this out, and make it, indeed, self-contradictory to say both "I see a reddish *x*" and "I see the black silhouette of *x*."

abnormal person explain to us, the way in which the basic action is performed (and this must be appreciated in the same way as the impossibility of explaining to a blind man what red literally looks like, or, if you wish, of our understanding what ultra-violet literally looks like). Suppose—just to take one case—a paralytic asks us what we do *first* when we raise an arm. We should be obliged to say we cannot answer, not because we do not know or understand what we do, but because we know and understand that there is *nothing* we do first. There is no series of steps we must run through, and since the request is implicitly for a *recipe*, we cannot say how we move our arm. A basic action is perfectly simple in the same sense in which the old “simple ideas” were said to be: they were not compounded out of anything more elementary than themselves, but were instead the ultimately simple elements out of which other ideas were compounded.

In one sense, then, we do, and in another we do not, know how we move an arm. But the sense in which we do not know is inappropriate. It is that sense which requires an *account*, and our incapacity for giving any such account is what has induced puzzlement, among philosophers and others, concerning the moving of an arm (and other basic acts generally). But this puzzlement should be dissipated upon the recognition that we have made a grammatical mistake in the inflected language of action. We have taken “moving an arm” as always a case of *C-1*, when *in fact C-4* is the standard case for normal persons moving normal arms normally. But having once committed this mistake, we look for a cause that is not there. And failing to find what we ought never to have expected to find, we complain that we do not know how we do move our arms. But of course we know. It is only that we cannot explain the manner of its doing. For there is no action, distinct from the action itself, to be put into the *explanans*. This is due to what I am terming the *givenness* of basic actions. Reference to basic actions belongs in the explanantia for explaining how things are done. So the paralytic, as long as he remains one, cannot understand: *Just raising the arm is what we do first*.

IV

A paralytic might think there is some *effort* he is not putting forth, by which, if he did or could put it forth, he might as a consequence move his arm. But I want to say that he cannot try to move his

arm if moving his arm is not already in his repertoire of basic actions. So in a sense he is right. If he could make the required effort, he could move his arm. But he cannot make that effort, cannot try, for he cannot in the only appropriate sense move his arm.

Consider the analogous situation with someone epistemically abnormal, say a deaf man. To ask a deaf man to try to hear a certain sound is rendered inappropriate by the fact that he is deaf. To try to hear, say, faint and distant music is to make an effortful listening. Only those who can already hear can make this effort. And what would count as trying (listening) in the deaf man's case? He could cup his ear, could place his ear to the ground, could contort his face and close his eyes. All this, however, is the pantomime of listening. Had he grinned or wagged a finger, it would have been as helpful. For there is no one thing that is better than any other in his situation. It is exactly this way with trying to move an arm. It is appropriate only to ask someone to try to move his arm when something externally inhibits normal movement, e.g., the arms are pinioned, and cannot be moved *freely* and *without effort*. But the paralytic cannot move his arm at all.

Consider these cases:

(a) I am a normal person who has swallowed a drug which gradually takes away the power to move an arm, rendering me, so long as it is in full effect, negatively abnormal. I make tests at five-minute intervals. It gets harder and harder to move my arm. And then I reach a point where I cannot move my arm and cannot *try* to. I have lost the power of trying, together with the power for doing.

(b) Someone thinks it would be spectacular to be able to extend and retract his fingernails, the way a cat does with its claws. We tell him it cannot be done, and he retorts that no one has ever tried, and he means to try. But in what should his trying consist? He could shake his fingers hard, could order them to extend, could pray, or could draw his soul up into a vast single wish. There is no rational way, for there is no way at all for a normal person. I don't mean that no one is or ever will be able to move his nails and to try to move them (e.g., with tight gloves on). If a man were prepared to suffer some sort of surgery, he might be able to cause his nails to go in and out, but we had not understood that he meant this by “trying.” It is after all not the way cats do it. It is more the way we move a loose tooth.

(c) I am a normal person, challenged to move

a normal stone. I take the challenge to imply the stone is not normal—perhaps it has some incredible density, or is fixed to a shaft driven deeply into the earth. But I decide to try, and the stone moves quite easily, having been a normal stone all along. So I conclude that the challenge was not normal. It turns out I was being asked to move the stone “the way I move my arm.” But this is not something I even can try to do. I can, with ridiculous ease, cause the stone to move. So I can try to cause it to move as well. But I cannot try to move it as a basic action—that would be a proper encounter with nothingness.

One can do with effort only what one can do effortlessly; and “trying,” the effort of will, is not something apart from the action that stands to it as cause to effect. It is the required action already being performed in untoward circumstances. Doing something with effort is not doing two things, any more than doing something gracefully is doing two things. Moving an arm is not then the result of an act of will: it *is* an act of will. But to speak of an act of will when the going is smooth is to behave a little like the dypsomaniac who wants to know what sorts of pink rats ordinary people see.⁴

It should be plain now why laughing, if performed as a basic action, is controversial. It is because whoever could so laugh would be positively abnormal, and we cannot understand what he does. In relation to him, we are in just the same position as the paralytic in relation to us. We lack a kind of gift.

V

It is easy enough to sympathize with those who feel an action is not intelligible unless we can find a causal picture for it. But this is only because they have taken intelligibility to consist in having a causal picture. Dominated by this requirement, they may tend to invent some such picture, populating their inner selves with entities whose job it is to serve the automotive functions demanded by the causal model of intelligibility. But I am asking that we do not strain, and that we use the causal model only where it is natural to use it.

That there are actions, like moving an arm, which do not really require any other action as

cause (and so no “inner” action as a cause) entails, I believe, no refutation of dualism. For all the distinctions I am thinking of are reproduced within the mental world, and cut across the distinction between body and mind. If, for instance, we take the description “*M* images *I*” where *I* is a mental image, then it is unclear, as it was in the case of “laughing” or “moves an arm,” whether *M* has performed an action or not, or, if an action, then a basic action or not. The whole declension works for, *C-1*: *M* may cause an image to appear in his mind, perhaps by taking a drug; *C-2*: Someone or something other than *M* may cause an image to appear in *M*’s mind; *C-3*: *M* is haunted by an image which appears spontaneously, recurrently, and unpredictably—a symptom, of perhaps a psychic disorder; and *C-4*: *M* simply produces an image, as I and all those with the requisite alpha rhythms are able to do, i.e., as a basic action.⁵

I shall not press for a full parity, though I *am* prepared to defend the view that there is a problem of Other Bodies precisely analogous to the problem of Other Minds. All I wish to emphasize is that, whatever disparities there may be between the concept of mind and the concept of body, men may be said to act mentally in much the same way that they may be said to act physically. Among the things I take Descartes to have meant when he said that we are not in our bodies the way a pilot is in a ship, is that we do not always do things, as pilots must with ships, by causing them to happen. We do not turn, as it were, an inner wheel in order, through some elaborate transmission of impulse, to cause an external rudder to shift and, by so doing, get our boat to turn. We act directly. But then neither am I in my *mind* the way a pilot is in a ship. Or rather, I sometimes cause things to happen with my body and with my mind, and I sometimes just act with them directly, as when I perform basic actions. It is best, however, to avoid similes. Any philosophical problems we have with ourselves would only reappear in connection with anything sufficiently similar to us to be a suitable analogue. But if we find ourselves unintelligible, nothing sufficiently similar to us to be helpful is likely to be more clear.

Columbia University

⁴ It is not difficult to see why it should be thought that there are two distinct things in the case of trying. It is because we often speak of trying and failing. So, if we can try and also succeed, trying is one thing and succeeding is another. And if succeeding consists in raising an arm, *trying* here must be something different, since failing consists in *not-raising* one’s arm, and trying then could hardly consist in raising it. But this is not the important sense of the word for the theory of action.

⁵ But I am not sure whether *we* are positively abnormal, or those who have no images are negatively abnormal.

VI. DESCARTES' VALIDATION OF REASON

HARRY G. FRANKFURT

IN the First Meditation, Descartes raises the possibility that there is a demon of unlimited power bent on deceiving those who reason. Later on in the *Meditations* he attempts to eliminate the doubts nourished by consideration of this possibility by developing proofs for the existence of God and by arguing that the benevolence of God guarantees the reliability of reason. To many critics, Descartes' procedure in this matter has seemed defective. They point out that his arguments rely upon the very rational faculty whose reliability is presumably at stake, and they insist that the attempt to validate reason is therefore vitiated by circularity.¹

In my opinion, those who have discussed this matter have often failed to understand Descartes' argument correctly. My purpose in this essay is to make clear just what question about reason Descartes found it necessary to ask in the *Meditations* and how he thought it possible to give a reasonable answer to it. When I have done so, I believe, it will be easier to judge fairly whether or not his reasoning is free of vicious circularity and other defects.

I

1. It may seem difficult to understand how anything can rationally be said in behalf of reason without transparently begging the question of whether reason is worthy of trust. Some commentators have consequently been attracted to the alternative that Descartes is not actually concerned with validating reason (the faculty by means of which, when we use it rightly, we perceive things

clearly and distinctly) at all, but with establishing the trustworthiness of memory.² Their position is that he makes no attempt to determine whether what is clearly and distinctly perceived (intuited) is properly to be regarded as true, but that he tries rather to provide grounds for trusting recollections of intuitions.³ Descartes is alleged by them to be preoccupied with the possibility that an omnipotent demon victimizes us by causing us to *think* we remember perceiving clearly and distinctly what we have never in fact intuited at all.

So far as I can see, this interpretation of the metaphysical doubt is inconsistent with Descartes' account of his doctrines and, in any case, it does not satisfactorily allow him to escape the charge of circularity.⁴ Descartes does indeed sometimes describe the problem with which he is concerned as one which may be encountered in contexts where something is remembered. But even in such contexts his problem is not to establish the reliability of memory. It is to validate propositions which are correctly remembered to have been intuited.

Suppose that two weeks ago, in the course of studying a geometry text, a person clearly and distinctly perceived that *p*. Suppose this person now correctly recalls having intuited on that earlier occasion that *p*. Given such a situation, Descartes wants to know whether the person is justified in accepting *p* as certainly true—whether, that is, the fact that he once intuited that *p* is now acceptable as conclusive evidence for *p*'s truth. His answer is that the person is justified if he knows that God exists, but not otherwise.⁵

¹ It is evident that Descartes is not guilty of circularity in the sense of offering an argument whose conclusion appears among its premisses. If his reasoning is circular at all, the circularity is, I take it, of a less formal variety. I shall not undertake to define this variety of circularity but shall assume that, in an intuitive way at least, its nature is sufficiently clear for the purposes of my essay.

² For an important defense of this alternative, cf. Willis Doney, "The Cartesian Circle," *Journal of the History of Ideas*, vol. 16 (1955), pp. 324-338.

³ The terms "to intuit" and "to perceive clearly and distinctly," and their corresponding derivatives, will be used interchangeably throughout this essay.

⁴ I have argued in support of these claims in "Memory and the Cartesian Circle," *Philosophical Review*, vol. 71 (1962), pp. 504-511.

⁵ "It is enough for us to remember that we have perceived something clearly, in order to be assured that it is true; but this would not suffice if we did not know that God exists and that He cannot be a deceiver." Charles Adam and Paul Tannery, eds., *Oeuvres de Descartes* (Paris, 1957), vol. VII, p. 246 (Latin); vol. IX, p. 190 (French); hereafter cited as "AT." Elizabeth

2. Before proceeding, it will be helpful to remove an obstacle which may otherwise stand in the way of a sound understanding of Descartes' problem. This obstacle is the erroneous notion that whenever Descartes says that something is indubitable, that is tantamount to his saying that it is true. Despite the fact that his metaphysical labors are largely devoted to exploring the relations between what is indubitable and what is true, it is not at all uncommon to find able writers on Descartes apparently overlooking the distinction between them.

For example, in the course of discussing the account given of mathematical propositions in the First Meditation, Leonard Miller remarks: "Descartes is puzzled by the nature of these propositions whose truth appears to be self-evident, for he is inclined to say both that we cannot possibly be mistaken about them provided that we apprehend them clearly and distinctly and that we can be deceived by the demon no matter how clearly and distinctly we perceive them."⁶ But what is the evidence that Descartes inclines to the view that even if there is a demon we cannot be mistaken about what we intuit? The only evidence cited by Miller is to the effect that what is intuited is not dubitable.⁷ As if in saying (as he wishes to do) that we cannot doubt what we are intuiting, Descartes is also saying (as he does not wish to do) that we cannot be in error about what we are intuiting but know it to be true whether or not we know that God exists.

Another capable critic, Willis Doney, ascribes to Descartes the settled opinion that even without knowing God's existence a person can know that what he at present clearly and distinctly perceives is true. In elucidating this interpretation, Doney

goes on to say: "Present clear and distinct perceptions were never subject to doubt. Anything so perceived did not depend on God as guarantor of its truth."⁸ Notice how readily Doney moves from speaking of something clearly and distinctly perceived as being not subject to doubt to speaking of it as being known to be true. Evidently he assumes that if something intuited is not subject to doubt, that is the same as its being known to be true. This explains why he is inclined to suppose that the second of the two statements just quoted from his essay is established when the first is shown to be true.

3. In fact, however, Descartes' metaphysical doubt is precisely a doubt whether being false is compatible with being indubitable. His position is that as long as the demon remains a possibility, we must acknowledge that what we intuit may be false. But he also holds that we cannot doubt the truth of what we intuit while we are perceiving it clearly and distinctly. "Our mind is of such a nature," he affirms, "that it cannot refuse to assent to what it apprehends clearly."⁹

While the intuition lasts, the inclination to believe what is being intuited is irresistible, and no doubt is then possible. But doubt may well arise at other times, if God's existence is unknown. "Before a man knows that God exists," Descartes declares, "he has an opportunity of doubting everything (viz., everything of which he does not have a clear perception present in his mind, as I have a number of times set forth)."¹⁰

Now at a time when we are intuiting nothing, we may recall having once perceived something clearly and distinctly. Descartes maintains that if we know that God exists, we are entitled to accept the fact that something was once intuited as con-

Haldane and G. R. T. Ross, eds., *The Philosophical Works of Descartes* (New York, 1955), vol. II, p. 115; hereafter cited as "HR." I shall at times depart from the text of HR in order to avoid inaccurate or clumsy translations.

⁶ "Descartes, Mathematics, and God," *Philosophical Review*, vol. 66 (1957), p. 452.

⁷ Cf. *ibid.*, pp. 451-452. I shall not consider whether more appropriate evidence is available for Miller's statement. My only point is that to show that Descartes believes it is impossible, demon or no demon, to *err* about what is intuited, it is not appropriate merely to offer evidence that he thinks it impossible, demon or no demon, to *doubt* what is intuited.

⁸ *Op. cit.*, pp. 325-326. As with the passage quoted from Miller, my point here is not that either of Doney's statements is false, but that the relation between them is not what Doney seems to suppose.

⁹ Letter to Regius (24 May 1640), AT, vol. III, p. 64. Descartes enunciates this doctrine on a number of occasions. Thus, in the Third Meditation, he says: "I cannot doubt that which the natural light causes me to believe to be true." AT, vol. VII, p. 38 (Latin); AT, vol. IX, p. 30 (French); HR, vol. I, p. 160. Again, in *Principles of Philosophy*, vol. I, p. 43, he says: "We are by nature so disposed to give our assent to things that we clearly perceive, that we cannot possibly doubt of their truth." It is essential that the questions raised by this doctrine be answered if Descartes' theory of knowledge is to be understood. For instance, what are his grounds for saying that it is impossible to doubt what is being intuited? Is his claim a contingent one, leaving the indubitability in question "merely psychological" and "subjective"? Or is there more to the claim than the assertion that as a matter of fact it happens that no one—no matter what he does—can experience any doubt about what he is intuiting? My own opinion is that there is more to it, but this is not the place to go as deeply into the matter as would be necessary to make it clear that this opinion is correct.

¹⁰ AT, vol. VII, p. 546; HR, vol. II, p. 333.

clusively establishing its truth; hence, the recollection then suffices to establish the truth of what we remember intuiting. But if God's existence is not known, he claims, we must acknowledge that what we remember intuiting may be false even though we once clearly and distinctly perceived it and were at that time incapable of doubting it.

For without the knowledge of God, "I can persuade myself of having been so constituted by nature that I can easily deceive myself even in those matters which I believe myself to apprehend with the greatest evidence and certainty."¹¹ Thus, our finding something to be indubitable—our apprehending it "with the greatest evidence and certainty"—cannot be regarded as itself a sufficient sign of truth. On the contrary, as long as we are ignorant of God's existence we must fear that it may be due to the malice of a demon who delights in making us find error irresistible.¹²

It is important to avoid the mistake of taking this to commit Descartes to the view that before God's existence is known any proposition can be doubted. As is suggested by the reservation introduced within parentheses in the passage cited above in footnote 10, Descartes carefully leaves open the possibility that there are propositions so simple that they cannot be thought of at all without being intuited. Such a proposition could never be doubted by anyone, with or without a knowledge of God's existence. For no one could doubt it without thinking of it, and anyone who thought of it would intuit it, and hence be unable to doubt it.¹³

4. To provide further clarification and support for my interpretation, let me discuss two important

passages in which Descartes tries to explain the problem he claims is solved by his demonstration of God's existence and veracity.

Toward the beginning of the Third Meditation, in the course of making clear why it is essential for him to inquire into the existence and nature of God, Descartes explains the metaphysical doubt to which he is still subject. He says that this doubt is aroused by considering the possibility that "perhaps some God might have endowed me with a nature such that I may be deceived even in respect of the things which seem to me the most manifest of all. . . . It is easy for Him, if He so wishes, to cause me to err even in those matters which I regard myself as intuiting . . . in the most evident manner."¹⁴ This seems to be a clear enough statement that Descartes is concerned with the possibility that even what is intuited may be false.

Some readers may nonetheless find it difficult to accept this understanding of Descartes' problem because of what he says immediately following the statement which has been quoted:

When I direct my attention to the things which I believe myself to be apprehending quite clearly, I am so persuaded of their truth that I cannot but break out into protestations such as these: Let me be deceived by whoever can do so, he will never be able to bring it about that . . . 2 and 3 could make more or less than 5; or that any other such things which I clearly see, can be other than I apprehend them as being.¹⁵

But there is surely no difficulty whatever in reconciling this statement with the view that Descartes is concerned with the possibility that

¹¹ AT, vol. VII, p. 70 (Latin); AT, vol. IX, p. 55 (French); HR, vol. I, p. 184.

¹² In the passage just quoted, Descartes speaks of matters which he *believes* himself to apprehend with great evidence and certainty, and not simply of matters which he apprehends thusly; and in passages quoted below, he speaks of things which *seem* to him most manifest, or matters which he *regards* himself as intuiting, and of things which he *believes* himself to be apprehending clearly. But this hardly jeopardizes the point that Descartes does not assume that whatever is indubitable is true. For if he did assume it, then whenever he believed himself to have intuited something it would be reasonable for him to believe that what he believed himself to have intuited is true. However, he says repeatedly that even when he does think that something has been intuited he must nonetheless acknowledge that it may be false (assuming that he does not know of God's existence). In any case, whatever doubts there may be about how to construe these passages are resolved by considering the latter part of the text cited in footnote 14, and the passages cited in footnotes 15 and 16.

¹³ Indeed, Descartes maintains explicitly that there are such propositions, and he regards the *cogito* as one of them (AT, vol. VII, pp. 145–146; AT, vol. IX, p. 114; HR, vol. II, p. 42). Now Descartes repeatedly asserts (e.g., in the passages I cite in footnotes 11 and 14), without any qualification or limitation whatever, that as long as he is ignorant of God's existence he must fear that a proposition may be false even though he intuit it in the most perfect way. He does not exempt the *cogito* from his general concern that unless God exists even what is intuited may be false. To be sure, the *cogito* is his paradigm of certainty, from which he derives the rule that whatever is intuited is true. But until this rule is vindicated the relation between the indubitability of the *cogito* and its truth is problematic. The *cogito* is so simple that it cannot be thought of without being intuited and found irresistible. But the fact that it can never be doubted is not identical with its being true or with its being known to be true. Descartes can still wonder whether its indubitability, however inescapable, is sufficient to establish its truth.

¹⁴ AT, vol. VII, p. 36 (Latin); AT, vol. IX, p. 28 (French); HR, vol. I, pp. 158–159.

¹⁵ *Ibid.*

even what is being intuited may be false. For Descartes clearly does not assert that no one could bring it about that he is deceived about the sum of 2 and 3 or about other things which he apprehends clearly. He only says that while apprehending them clearly, he is "so persuaded of their truth" that he cannot help protesting that he cannot be deceived about them.

Now the fact that he is persuaded of their truth to this extent is not the same as their being true; nor is his inability at the time to conceive that he could be mistaken the same as his being in fact free of error. In this passage, Descartes describes the convictions he is irresistibly inclined to hold under certain circumstances, and he reports the assertions he feels urgently moved under these circumstances to make. But he does not say either that the convictions are reasonable or that the assertions are true.

The second passage I wish to discuss is from Descartes' "Reply" to the second set of "Objections" against the *Meditations*:¹⁶

There are other things which our understanding also perceives very clearly, when we pay close attention to the reasons on which our knowledge of them depends; and while we are doing so we cannot doubt them. But since we can forget those reasons and yet remember the conclusions which were drawn from them, the question arises if we can have a firm and immutable conviction concerning these conclusions during the time we recollect that they were deduced from principles which were most evident; for this recollection must be supposed in order that they may be called conclusions. My answer is that such conviction can be had by those who, in virtue of their knowledge of God, are aware that the faculty of understanding given by Him must tend toward truth; but others cannot have it.

Here Descartes is supposing that someone once deduced a conclusion from premisses which he was at the time intuiting, but that he no longer remembers these premisses. Thus, the person does not now perceive clearly and distinctly that the conclusion follows from premisses which he is intuiting. He only remembers that the premisses were evident to him at one time—i.e., that he once perceived them clearly and distinctly—and that he deduced the conclusion from them—i.e., once intuited that from them the conclusion follows.

While he was intuiting the premisses, Descartes maintains, he was not able to doubt them; nor was he able to doubt that the conclusion follows

from the premisses while he was intuiting its relation to them. But now he is free to doubt these things, and his problem is to decide whether doubt is justified or whether what he remembers suffices to establish the soundness of the argument being considered. The question is this: given that a proposition has been intuited to follow from premisses which were themselves intuited, is it possible that the proposition should be false? Is it possible, in other words, that a proposition should be perceived clearly and distinctly to follow from a set of premisses—i.e., be deduced from the set—without actually following from it? And is it possible that premisses should be evident—i.e., be perceived so clearly and distinctly as to be subject at the time to no doubt at all—without being true?

5. The metaphysical doubt arises for Descartes when he remembers some intuition, but it is not a doubt about the reliability of memory. Indeed, it should now be apparent that there is no reason why metaphysical doubt may not arise even in situations in which there is no recollection of anything being intuited. For what is required as a context for metaphysical doubt is not necessarily a situation in which a person recollects having intuited something. A suitable context is provided by any set of circumstances in which a person can consider the validity of an intuition.

Thus, suppose that at a certain time one man *A* perceives something clearly and distinctly, and that another man *B* knows right then and there that *A* is doing so. Suppose further that *B* is uncertain whether the occurrence of *A*'s clear and distinct perception is sufficient to establish the truth of what is being intuited by *A*. Then *B* is engaging in metaphysical doubt about what *A* is intuiting, and it is quite evident that *B* need not be remembering anything at all while doing so.

Descartes' failure to make this altogether clear is rather easily explained. When he discusses these matters in the *Meditations*, the development of his metaphysics is at a stage in which he does not know that anyone exists but himself. He does not, accordingly, consider any intuitions but his own. Since he cannot doubt the validity of his own intuitions while they are occurring, he can engage in metaphysical doubt about them only after they have occurred and while he recollects their occurrence. Thus, metaphysical doubt arises for Descartes only when he recollects having perceived something clearly and distinctly. But this is due to the order in which matters are taken up in the

¹⁶ AT, vol. VII, p. 146 (Latin); AT, vol. IX, pp. 114-115 (French); HR, vol. II, pp. 42-43.

Meditations and not to the nature of the doubt itself.

II

6. Assuming that the preceding interpretation of Descartes is correct and that he is trying to validate reason (intuition) by coping with metaphysical doubts concerning the truth of what is intuited, can his procedure in the *Meditations* escape the common charge of circularity?

One of the first to be struck by the apparent circularity of Descartes' reasoning was Arnauld, who made the point as follows:

The only secure reason we have for believing that what we clearly and distinctly perceive is true, is the fact that God exists. But we can be sure that God exists only because we clearly and evidently perceive it. Therefore, prior to being certain that God exists, we should be certain that whatever we clearly and evidently perceive is true.¹⁷

Now if to be sure of something were to be unable to doubt it, then Arnauld would be mistaken in supposing that, according to Descartes, we can be sure of God's existence only if we already know that whatever is intuited is true. For if we intuit that God's existence follows from premisses which are at the same time also intuited, then while these intuitions occur we shall be unable to doubt that God exists even if we do not know that whatever is clearly and distinctly perceived is true. We shall be sure of God's existence during that time because we shall actually be intuiting that God exists and hence we shall be, as we are whenever we are intuiting something, irresistibly impelled to believe it.

As a matter of fact, Descartes does believe that all the steps in the proof of God's existence can be intuited simultaneously.¹⁸ It is not implausible (and I shall suppose) that he also thinks it possible to intuit simultaneously not only these steps but also the further steps involved in arguing that the truth of whatever is intuited is guaranteed by the existence of God. Thus, without begging any questions or in any way committing the fallacy of circularity, Descartes allows the possibility of our being sure that whatever is intuited is true. For our belief in this principle may be rooted in present intuitions so that we are incapable of doubting it no matter what else we know or believe.

7. Of course, this hardly settles whether or not Descartes argues in a circle. What it shows is merely that without relying upon a circular argument it is possible, in terms of Descartes' position, to *be sure* that whatever is intuited is true. To eliminate doubt and attain assurance of this one need only run through the argument that God exists and validates reason, keeping all relevant intuitions in mind at once. But being sure of the principle that what is intuited is true is not the same as knowing it to be true, and it would certainly seem that—like other objects of intuition—this principle can well be doubted when one is not intuiting it but is, say, only remembering that it has been intuited.

The remarkable thing, however, is that Descartes denies this. Indeed, he maintains quite straightforwardly that after it has once been demonstrated that what is intuited is true, one need not run through all the intuitions comprising this demonstration each time it is necessary to invoke the divine guarantee of the truth of what is clearly and distinctly perceived. What he says is this:

After I have recognised that there is a God . . . and have inferred that what I perceive clearly and distinctly cannot fail to be true, although I no longer pay attention to the reasons for which I judged this to be true, provided that I recollect having clearly and distinctly perceived it, no contrary reason can be brought forward which could ever cause me to doubt its truth; and thus I have a true and certain knowledge of it.¹⁹

When one wishes to invoke the principle that what is intuited is true, Descartes assures us, it is sufficient to remember having demonstrated it. It is not necessary to repeat the intuitions comprising its demonstration.

But why not? Why should this principle be established by the recollection that it was once demonstrated when, in general, recalling that something has been intuited is not sufficient to establish it? Does it not surely seem that Descartes is guilty here of the egregious blunder with which he has so often been charged? For does he not sanction accepting as evidence for the principle that intuitions are true the fact that this principle was once intuited? And would not anyone who accepted such evidence for the principle be begging the entire question—the question, precisely, of whether such evidence is acceptable?

¹⁷ AT, vol. VII, p. 214 (Latin); AT, vol. IX, p. 166 (French); HR, vol. II, p. 92.

¹⁸ AT, vol. V, pp. 148–149; Charles Adam, ed., *Entretien avec Burman* (Paris, 1937), pp. 9–13.

¹⁹ AT, vol. VII, p. 70 (Latin); AT, vol. IX, pp. 55–56 (French); HR, vol. I, p. 184.

8. Before taking answers to these questions for granted, let us examine carefully the last quoted passage, in which Descartes explains how things stand, in his view, when the existence of God has once been demonstrated and when it has once been seen clearly and distinctly that from this it follows that what is intuited is true.

Notice what Descartes claims to be the case when he recollects having intuited that God guarantees the truth of what is clearly and distinctly perceived. He claims that then "no contrary reason can be brought forward which could ever cause me to doubt its truth." He does not assert that when he recollects having intuited that the principle in question is true he cannot then *experience* doubts as to its truth. Nor does he deny what is in any case surely not deniable—that he can always *state* that he doubts it. But he indicates that any such statement, at least from the logical point of view, will be capricious. For he cannot, Descartes claims, *have a reason* for doubting it.

Now the possibility that there is not a veracious God is, of course, accepted by Descartes as a reason for doubting that whatever is intuited is true. So his claim involves the view that when one remembers both that God's existence was once intuited and that it was also intuited at one time that what is intuited is true, it is not then reasonable to entertain the possibility that a veracious God does not exist. But why is it unreasonable? We may doubt other things which we recall having intuited. Why can we not with equal reason doubt the existence of a veracious God when we remember intuiting it?

Consider just what it is that is being recalled in the case at issue—namely, that exercising reason in the most rigorous way (i.e., accepting only what is clearly and distinctly perceived) results in the intuition that a veracious God exists. That this is what results when reason is used in the most impeccable manner means that the soundest use of reason leads to the exclusion of the possibility that there is an omnipotent demon and, indeed, to the exclusion of the possibility that man's being derives from any source lacking in power or in perfection. Descartes has undertaken to show that intuition provides no basis for supposing that what is

intuited may be false, and it is the establishment of this conclusion that is recalled.

Far from leading to the discovery of reasons for mistrusting reason, Descartes attempts to show, the most conscientious use of reason leads to the discovery that such mistrust has no rational ground. Hence, when someone remembers having perceived clearly and distinctly that intuition is guaranteed by God, what he remembers is that there is no good reason for doubting the trustworthiness of intuition. In other words, he remembers something which makes it plain that the metaphysical doubt is utterly capricious.²⁰

9. How Descartes' reasoning is to be understood and evaluated will become more apparent if the focus of this discussion is broadened to include the general nature of the enterprise he undertakes in the *Meditations*. As everyone knows, Descartes is largely concerned in the *Meditations* with the problem of scepticism. Now so far as scepticism with regard to reason is concerned, the classical gambit of the sceptic is to show that the use of reason leads ineluctably to the conclusion that reason is unreliable. Indeed, this is the sceptic's only available gambit, if he is to argue at all. All he can do is attempt to provide arguments which demonstrate the untrustworthiness of reason.

Naturally, his attempt will only succeed if his arguments are good ones—i.e., if he can give good reasons for regarding with suspicion the significance of good reasons. The sceptic must show that reason can be turned against itself, by showing that there are reasons of the very strongest sort for doubting the reliability of reason. We may say, then, that the sceptic's arguments are designed to provide a *reductio ad absurdum* of the assumption that reason is reliable.

In order to dispose of scepticism with regard to reason, therefore, Descartes believes he need only show that the sceptic's attempt to overthrow reason is a failure. And he regards this as having been accomplished as soon as he shows that the most rigorous use of reason does not lead to a mistrust of reason but, rather, to conclusions which exclude all basis for such mistrust. What Descartes takes to be his task, in other words, is to show that the sceptic's *reductio* argument cannot be generated.

²⁰ That Descartes regards this (rightly or wrongly) as sufficient to establish the truth of what is intuited is readily apparent in the latter part of the passage cited in footnote 18, and in such passages as this one from his "Reply" to the second set of "Objections": "After becoming aware that God exists, it is necessary to imagine that He might be a deceiver if we wish to cast doubt on what we perceive clearly and distinctly; and since we cannot even imagine that He is a deceiver, we must admit these things as most true and most certain." AT, vol. IX, p. 113; HR, vol. II, p. 41. The emphasis has been added by me to call attention to the "negative" character of Descartes' procedure: he establishes truths by removing the grounds for doubting them rather than by proving them more directly to be true.

He attempts to do this by offering a proof that there is an omnipotent deity who is not a deceiver and whose existence, accordingly, entails that reason is reliable. The value of this proof depends on its success in showing where the right use of reason in fact leads: neither to the conclusion that there is an omnipotent demon devoted to deception, nor to any other conclusion involving the untrustworthiness of reason. The proof purportedly makes it clear that when reason is put properly to use it produces reasons of the very best sort (i.e., clear and distinct ideas) for trusting reason. It produces no such reasons for mistrusting reason, and so the sceptic's attempt to reduce reliance on reason to absurdity is seen to fail.

Descartes' argument is thus to be understood as an attempt to show that there are no good reasons for believing that reason is unreliable. Its purpose is to reveal that the hypothesis which provides a basis for mistrusting reason is not one which reason supports, and that the mistrust of reason must accordingly be regarded as irrational. If reason is properly employed—that is, if we give assent only to what is intuited—we are not led to doubt that reason is reliable. On the contrary, we are led to assent to the propositions that God exists and that He guarantees the reliability of reason.

As long as the existence of an omnipotent demon had to be acknowledged by reason to be a possibility, it had to be acknowledged that the use of reason might lead to the conclusion that the demon does exist and that therefore reason is not reliable. Hence, what is essential in Descartes' argument is not so much the discovery of the existence of a benign deity, but the discovery that reason leads to the conclusion that such a deity exists.²¹

10. Suppose someone recalls having perceived

something clearly and distinctly and wonders if he is entitled to regard what he intuited as certainly true. If he does not know whether or not the sceptic can succeed in the attempt to provide a *reductio ad absurdum* of the trustworthiness of intuition, then he must properly be uncertain whether whatever is intuited is true. For all he knows, it may be possible to find impeccable grounds for regarding reason as unreliable—for example, by showing clearly and distinctly that there is an omnipotent demon bent on spoiling the work of reason.

But such doubts are legitimately dispelled, Descartes maintains, if the person can recall that the existence of a veracious God has been demonstrated, for he then recalls that reason does not fall victim to the sceptic's *reductio* but instead decisively escapes it. That the existence of a veracious God has been clearly and distinctly perceived answers the question concerning the possible success of the sceptic's line of argument. It means that the sceptic's line of argument fails. This question being answered, there remain no reasonable grounds upon which to base metaphysical doubts.

It is evident that Descartes' argument does not suffer from the circularity with which it is commonly charged.²² Metaphysical doubt concerns the truth of what is intuited, and the removal of this doubt is effected without assuming that what is intuited is true. It is removed simply by the knowledge that a certain demonstration has been successfully accomplished. This knowledge is, of course, the knowledge that certain things have been clearly and distinctly perceived. But it is not required that the *truth* of these things be supposed, and so the question is not begged. All that is relevant to the removal of metaphysical doubt is that the sceptic's *reductio* be discovered not to

²¹ Alan Gewirth makes a similar point in his excellent essay, "The Cartesian Circle," *Philosophical Review*, vol. 50 (1941), pp. 389–390: "The ground upon which the clear and distinct perception of God's existence and veracity is regarded by Descartes as overthrowing the metaphysical doubt, then, is that the rationality of the former reveals the 'reasons' of the latter to be irrational." Gewirth's work on Descartes, to which I am very greatly indebted, deserves more attention. In my opinion, his three essays on Descartes' theory of knowledge are by far the best things of their kind in English. The other two essays are "Experience and the Non-Mathematical in the Cartesian Method," *Journal of the History of Ideas*, vol. 2 (1941), pp. 185 ff; and "Clearness and Distinctness in Descartes," *Philosophy*, vol. 18 (1942), pp. 17 ff.

²² But Descartes' reasoning may well be defective, and it may even be circular. Indeed, the following serious question must be raised about it. Given that reason leads to the conclusion that reason is reliable because a veracious God exists, may it not also lead to the conclusion that there is an omnipotent demon whose existence renders reason unreliable? Of course, these conclusions are incompatible, and if the proper use of reason established both of them it would mean that reason is unreliable. But surely Descartes cannot take for granted that this is not the case. His procedure does, therefore, seem to beg the question, though in a rather different way than has generally been thought.

To put the same point differently: what Descartes attempts to do is to provide a proof of the consistency of reason; but this proof is decisive only if we suppose (thereby begging the question?) that reason is consistent; since otherwise it might still be possible to construct an equally cogent proof of the inconsistency of reason. Descartes seems to have been unaware that his procedure is open to this line of criticism, and I do not propose to enter here into the difficult investigation that would be necessary in order to arrive at a just evaluation of its significance.

materialize, and this discovery can be made and recalled without anything intuited being supposed to be true.

11. If I am correct in what I have ascribed to Descartes, his reasoning in the *Meditations* is designed not so much to prove that what is intuited is true as to show that there are no reasonable grounds for doubting this. Now it may be objected that in that case he leaves the main question still open, since it may be that what we intuit is sometimes false even if we can have no reasonable grounds for supposing so. Whatever may be the weight of this objection, it bears against Descartes' doctrines and not against the authenticity of my interpretation of them. Indeed, some confirmation for my interpretation is to be seen in the fact that Descartes acknowledges that an objection of this sort may be raised against his position.

Thus, he begins a summary statement of his position by making it clear that in his view "if . . . we can never have any reason to doubt that of the truth of which we have persuaded ourselves, there is nothing more to inquire about; we have all the certainty that can reasonably be desired."²³

Immediately thereafter, he anticipates the objection that certainty based upon the unavailability of reasonable grounds for doubt is compatible with the falsity of that of which there is certainty. It is particularly interesting to consider his manner of formulating this objection and of responding to it. Concerning something of which

we have "all the certainty that can reasonably be desired," he says:

What is it to us if someone should feign that the very thing of whose truth we are so firmly persuaded appears false to the eyes of God or of the Angels and that hence, speaking absolutely, it is false? Why should we concern ourselves with this absolute falsity, since we by no means believe in it or even have the least suspicion of it? For we are supposing a belief or a conviction so strong that nothing can remove it, and this conviction is in every respect the same as perfect certitude.

Evidently Descartes recognizes his position to entail that from the fact that we know something with perfect certitude, it does not follow that it is, "absolutely speaking," true. He concedes, that is, that he has not proven that whatever is intuited is absolutely true.

If what is perfectly certain may be absolutely false, Descartes suggests, the notions of absolute truth and absolute falsity are irrelevant to the purposes of inquiry. Presumably he would wish them to be replaced with other notions of truth and falsity. But what are these notions and how are they related to those which Descartes rejects? This line of reflection leads rapidly to a large number of questions: concerning Descartes' conceptions of certainty, of knowledge, and of the relation between knowledge and reality. To explore these matters is likely not only to enhance the understanding of his position, but to be of considerable philosophical interest as well.²⁴

The Rockefeller Institute

²³ Both this passage and the one quoted next are from Descartes' "Reply" to the second set of "Objections," AT, vol. VII, p. 145 (Latin); AT, vol. IX, pp. 113-114 (French); HR, vol. II, p. 41.

²⁴ For helpful comments on an earlier version of this paper, I wish to express my appreciation to Carlos Blanco, Bruce Lercher, Maurice Mandelbaum, C. Wade Savage, Sydney Shoemaker, and Richard Sorabji.

VII. AUGUSTINE ON SPEAKING FROM MEMORY*

GARETH B. MATTHEWS

IN the twelfth chapter of his little dialogue¹, *De Magistro*, Augustine makes a most peculiar claim regarding memory:

When a question arises not about what we sense before us, but about what we have sensed in the past, then we do not speak of the things themselves, but of images impressed from them on the mind and committed to memory.²

Augustine seems to be saying that whenever we are asked about familiar, but absent, sensible things we respond by changing the subject, that is, by talking of our memory images instead. Now it is certainly true that we sometimes do this sort of thing. When asked to describe something from memory I may plead that I have only the haziest (mental) picture of that particular thing. I may offer this as a reason for not trying to answer the question itself. Or, again, I may preface my answer to the question with the claim that I have an especially clear (mental) picture of the object enquired about. I may use such a prefatory claim to give authority to my report and elicit credence in it. But these two responses would lose their point if we were never able to describe the absent things themselves. For admission that one has only the haziest recollection of something would not really count as a reason for not trying to describe that thing unless we were sometimes able to succeed in such descriptions. And claiming to have a very good recollection of something gives weight to one's words simply because a good recollection affords a good basis for making assertions about the things themselves.

Augustine might more plausibly have said that we speak from memory of sensible things *according to our memory images*. But he did not. What he did say is so peculiar that one cannot help wondering whether he really meant it.

I

We might hope to eliminate or at least mitigate the paradox by finding a better translation for Augustine's Latin. Perhaps "*loquimur*" in the context "*non iam res ipsas, sed imagines . . . loquimur*" should not be rendered "speak of," as I have rendered it, but in some quite different way.

I have checked a half-dozen standard translations of this passage. Two of these do seem to relieve the paradox. This is one of them:

. . . ce ne sont plus les choses elles-mêmes que nos paroles indiquent, mais les images. . . .³

The doctrine expressed here is that some of our words point to (or indicate) things which are present, whereas others point to (or indicate, or perhaps are signs of) our memory images rather than the things themselves. Unfortunately this translator pays a very high price for his version; he has to change the subject from "we" to "our words." The result is more of a gloss than a translation.

The other divergent translation is this one:

But when a question is asked not regarding things which we perceive while they are present, but regarding things of which we had sense knowledge in the past, then we express in speech, not the realities

* An earlier version of this paper was read at the Conference on Medieval Studies at Western Michigan University on March 6, 1964.

¹ Sometimes Augustine uses "*memoria*" very broadly to include not only memory but also imagination and conception generally (cf. especially Book X of the *Confessions*). The vast Augustinian secondary literature includes many good discussions of this concept of *memoria* (e.g., Etienne Gilson, *The Christian Philosophy of St. Augustine* [New York, 1960], especially pp. 75 and 100-104). My subject here is something more limited. I mean to be talking about what we, as well as Augustine, would call memory claims. So far as I can tell, what Augustine says specifically about these has been neglected in the secondary literature.

² "Cum vero non de his, quae coram sentimus, sed de his, quae aliquando sensimus, quacritur, non iam res ipsas, sed imagines ab eis impressas memoriaeque mandatas loquimur" (*De Magistro* 12.39). I follow the critical edition of *De Magistro* (*Corpus Scriptorum Ecclesiasticorum Latinorum*, vol. 77, sect. 6, pars 4 [Vienna, 1961], p. 48). However, the standard Maurist or Benedictine edition (J. P. Migne, *Patrologia Latina*, vol. 32 [Paris, 1845], col. 1216) differs here only inessentially in punctuation and in pronominal usage.

³ *Oeuvres de Saint Augustin* (Paris, Desclée de Brouwer et Cie, 1941), vol. VI, p. 105. The translation is by F. J. Thonnard.

themselves, but the images impressed by them on the mind and committed to memory.⁴

This translator uses "express in speech" for "*loquor*"; he reserves "speak about" for "*dico*." And he renders "*res*" as "reality," reserving "thing" or "things" for appropriate Latin pronouns (e.g., he translates "*de his, quae . . . sentimus*" as "of things which we perceive").

The force of this translation is to suggest that, according to Augustine, we can do something called "expressing realities in speech" and something called "expressing in speech images impressed by realities on the mind and committed to memory" and that these activities are distinguishable from merely speaking about things and speaking about the images of things. Such a suggestion is welcome, insofar as it gives us hope that Augustine is not after all denying what seems so obviously true, viz., that we do sometimes speak of sensible things from memory.

But what are we to make of the suggestion? Just what is it to express realities in speech? And how does this differ from speaking of (or about) things? I would not know how to explain, let alone defend, such a distinction, either in itself, or as an interpretation of Augustine. The translator provides no explanatory footnote. In the absence of satisfactory clarification, I conclude that the relief from paradox which this second translation promises is delusory.

Moreover, I have no helpful alternative translation to offer. This leads me to take seriously the possibility that Augustine in fact said what I have translated him as saying. Does the surrounding discussion in *De Magistro* make it plausible for one to conclude that Augustine really did say this and mean what he said?

II

I think it does.

Augustine's main concern in *De Magistro* is to say how it is we succeed in learning things.⁵ But in the course of the discussion he takes up several subsidiary matters, including a worry about how one person can succeed in answering another's question. "If you had entertained that very first

question of mine solely according to the sound of its syllables," says Augustine to his interlocutor, Adeodatus, "you would have given me no answer, for I should have seemed to you to ask nothing." To this Adeodatus replies, "I agree with you that we cannot converse at all unless upon hearing words our mind is led to the things of which the words are signs."⁶ That is, one man cannot succeed in responding to another man's speech unless the mind of the second man is led to the things signified by the words of the first man.

Of course Adeodatus' comment engenders its own puzzlement. How can one man's mind be led to the things signified by the words of another man? The *De Magistro* answer to this last question is that the first man's words prompt the second man to look at the things signified.⁷ Without pushing the regress of puzzlement further, let us consider the significance of this last Augustinian doctrine.

For a person to be able to look at the things which certain words signify, those things must be available to him for inspection. According to Augustine "intellectual things" (*intelligibilia*) are available to every man according to his inner light.⁸ But "sensible things" (*sensibilia*) are directly available only to one who happens to be located in their immediate vicinity. So when talk turns to familiar, but absent, sensible things, it cannot be the things themselves one's mind is prompted to look at; it must be images of them "impressed on the mind and committed to memory."

If we grant Augustine the natural supplementary principle that that which a man's words signify is what he speaks of, we arrive at the paradox with which we began, viz.:

When a question arises, not about what we sense before us, but about what we have sensed in the past, then we do not speak of the things themselves, but of images impressed from them on the mind and committed to memory. Indeed I do not know how we come to call (the things we speak of) real, since what we look at are counterfeits (i.e., images), unless it is because we explain, not that we see and sense them, but that we have seen and have sensed them. Thus we carry these images in the recesses of the memory as proofs (*documenta*) of things sensed before. Contemplating them in the mind we tell no falsehood when we

⁴ *The Greatness of the Soul* (and) *The Teacher* (Westminster, Md., Newman, 1950), pp. 178-179. The translation is by Joseph M. Colleran.

⁵ See especially *De Magistro* 11. 36-8 and 14.45-6.

⁶ *Ibid.*, 8.22.

⁷ *Ibid.*, 8.23, 10.33-5, 11.36.

⁸ *Ibid.*, 12.40.

speaking in good conscience, but indeed they are for us proofs (*documenta*).⁹

Augustine is maintaining that memory claims of a certain sort are (so to speak) ellipses. Questioned about what color my neighbor's house is I answer from memory, "The house is brown." But according to Augustine such a response is (as we might say) elliptical. To express myself completely I should have to say, "I have the impression that the house is brown," or "It is my recollection that the house is brown."

Augustine's talk of memory images as "proofs" (*documenta*)¹⁰ of things sensed before might suggest that he considers memory images an infallible guide to the truth about absent sensible things. However, the explicit assurance that he gives us is this: "Contemplating (these images) in the mind we tell no falsehood when we speak in good conscience. . . ." Augustine does not say that when we speak in good conscience we *make no mistake* about the absent things; he says only that *we tell no falsehood* (*non mentimur*). I take it that what Augustine has in mind here is the fact that a person cannot be mistaken about his own impressions in the straightforward way he can be mistaken about sensible things themselves. Of course one may lie about one's impressions. Or by slip of the tongue or through ignorance of the language one may misdescribe them. But the impressions, unlike the things themselves, simply are what they seem to be.

As we noted earlier, Augustine seems to suppose that a person trying to answer a question is limited to reporting on what he can in some way perceive. Thus a person "looking at" memory images, rather than at things themselves, is limited to reporting on those images. If then images simply are what they seem to be, memory images do "prove" or establish the answers offered by a candid and conscientious respondent.

The moral of all this seems to be that when answering questions about sensible things from memory we need worry only about the conscientiousness of our report; we are unable to give the kind of answer that could be honestly mistaken (i.e., the kind of answer that would be about the things themselves rather than about the respondent's own impressions).

III

What of it? Suppose Augustine did say and mean to say that when we answer questions about sensible things from memory we really speak, not of the things themselves, but of our memory images. Why is this significant?

I have suggested that Augustine is led to make this paradoxical claim upon considering how we manage to answer questions in general and, in particular, how we can succeed in answering questions about sensible things from memory. Augustine decides that we are able to do the latter by looking at memory images the questioner's words remind us of. And that leads him to suppose that, in making such answers, we speak of the images we look at rather than of the things themselves (which, of course, we can no longer see). But we are asked about sensible things and we answer by speaking of our images of them. So we change the subject. Indeed, Augustine seems to think we have to change the subject. Thus, after beginning with a worry about how it is we succeed in answering questions about sensible things from memory, Augustine ends with the unwelcome assurance that in fact we do not succeed.

In this way Augustine's analysis miscarries. The analysis seems called for by the puzzling character of a common accomplishment. But from the analysis it follows that there really is no such accomplishment. So what we are left with is, at best, an explanation of how it is that we seem to be able to do something we really cannot do.

I suggest that the reason the analysis miscarries is that Augustine is overly restrictive in construing the puzzle he is dealing with. "How can we answer questions about sensible things from memory?" can be taken in at least two rather different ways, and these ways ought to be distinguished. Taken in one way, the question means: (1) What mental mechanism makes it possible for us to speak of things from memory? Taken in another way, it means: (2) How does one make a response count as answering a question about absent sensible things? Augustine worries about (1) to the neglect of (2); and for this reason his analysis of speaking from memory miscarries.

Augustine's supposal that we answer questions about sensible things from memory by looking at

⁹ *Ibid.*, 12.39.

¹⁰ "*Documentum*" from "*docere*," to teach, also carries the suggestion of instruction. This connotation, though important for the context in which the passage quoted above appears, is inessential to the quotation itself.

our memory images is certainly an appropriate approach to (1). It may or may not be true that some (or all) people need to "look at" their memory images to be able to answer such questions. But, true or not, there is at least nothing paradoxical or unreasonable about supposing this to be the case.

Question (2), however, requires a different approach. It will not be satisfactory to say that a response based upon a look at one's memory images automatically counts as answering a question about absent sensible things. What the respondent says must be made liable to confirmation and disconfirmation by evidence other than the "proof" of his own memory images if it is to count as an answer to a question about absent sensible things (and not merely as a report of the respondent's own impressions). In fact, an answer considered confirmable by the evidence of other people's memory reports (*inter alia*) is by that very

fact not about impressions but about the things themselves, whatever the mechanism of each individual's recall may be. And this is the beginning of a proper answer to (2).

The miscarriage of Augustine's analysis and the paradox in which it results are instructive because they point up the danger in worrying about the mechanism of question answering to the neglect of the logic of enquiry. Augustine is right in thinking that there is no mental mechanism which can give us direct access to absent sensible things. But he is wrong in concluding from this that, when talk turns to such things, we are therefore limited to giving introspective reports on our mental images. In fact the situation is quite otherwise. By making our statements liable to correction from other sources we overcome the imagined limits of mental mechanism and manage to answer questions about the absent things themselves. We need not change the subject unless we want to.

University of Minnesota

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston	James M. Edie	Richard H. Popkin
Alan R. Anderson	Peter Thomas Geach	Wesley C. Salmon
Kurt Baier	Adolf Grünbaum	George A. Schrader
Richard B. Brandt	Carl G. Hempel	Wilfrid Sellars
Lewis W. Beck	Jaakko Hintikka	J. J. C. Smart
Roderick M. Chisholm	Raymond Klibansky	Wolfgang Stegmüller
L. Jonathan Cohen	Benson Mates	Manley H. Thompson, Jr.
James Collins	John A. Passmore	G. H. von Wright
Michael Dummett	Günther Patzig	John W. Yolton

VOLUME 2/NUMBER 3

JULY 1965

CONTENTS

I. ROBERT E. BUTTS: <i>Necessary Truth in Whewell's Theory of Science</i> . . .	161	V. MAURICE MANDELBAUM: <i>Family Resemblances and Generalization Concerning the Arts</i> . . .	219
II. JOSEPH MARGOLIS: <i>Recent Work in Aesthetics</i> . . .	182	VI. ALASTAIR MCKINNON: <i>Unfalsifiability and the Uses of Religious Language</i> .	229
III. PETER ACHINSTEIN: <i>The Problem of Theoretical Terms</i> . . .	193	VII. KEITH CAMPBELL: <i>Family Resemblance Predicates</i> . . .	238
IV. GERHARD GENTZEN: <i>Investigations into Logical Deduction: II</i> . . .	204		

UNIVERSITY OF PITTSBURGH PRESS

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be type-written with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

REPRINTS

Authors who are subscribers will receive 50 reprints gratis. Additional reprints can be purchased through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues are sold at the rate of two dollars to individuals, and three dollars to institutions. Correspondence regarding subscription and back orders may be addressed directly to the publisher (University of Pittsburgh Press, University of Pittsburgh, Pittsburgh, Pennsylvania 15213).

* * *

The *American Philosophical Quarterly* is published quarterly in January, April, July, and October by the University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, Pennsylvania, 15213.

Second-class postage paid at Pittsburgh, Pennsylvania.



I. NECESSARY TRUTH IN WHEWELL'S THEORY OF SCIENCE

ROBERT E. BUTTS*

I. INTRODUCTION

WILLIAM WHEWELL'S fifty-year long career as scientist and historian and philosopher of science was in many ways a perverse one. In keeping with his training, his interests, and the most influential intellectual trends of his age, Whewell should have been a philosophical empiricist.¹ His own work in science was for the most part of the most narrow empirical kind. He collected and classified minerals, attempted to measure the density of the earth at the bottom of Dolcoath coal mine shaft, made a monumental descriptive study of the tides off the coast of England, invented an instrument, called an "anemometer," for measuring the force and direction of the wind, and prepared detailed notes on the architecture of German churches. The tenor of the intellectual times in Britain was also clearly empirical: the utilitarianism of John Stuart Mill rose to popularity; in Scotland Dougald Stewart continued to develop the fortunes of Scottish common sense empiricism; William Hamilton loudly proclaimed the virtues of his variety of empiricism; and even in Cambridge great friends of Whewell—like Sir John Herschel—championed a different variation on empiricism. Yet, in spite of what his own scientific work and the philosophical tendencies of his age might have suggested to him, Whewell's philosophy of science emerged as one of the last great rationalist systems, complete with a metaphysics, a theology, and a theory of morals. In its finished or nearly finished form, Whewell's

philosophy was, in Britain at least, an anomaly. For his system was, in its basic features and in many of its details, much more like the systems of Leibniz and other seventeenth-century rationalists than like those of his British contemporaries.²

The dimensions of Whewell's perversity loom even larger when one takes into account the fact that though Newton and Bacon were his announced great scientific and philosophical heroes, the most fundamental features of his philosophy derive, in spirit if not always in detail, from his study of Plato and Kant. He translated Plato's dialogues into English, and always had a warm spot for his theory of ideas, though he did not hesitate to criticize some of its implications and limitations. The Kantian motives of his philosophy of science have not yet been clearly worked out by Whewell's commentators, but the fact seems to me to be that in curious ways Whewell's mature system owes more to a basic understanding of both the merits and the limitations of Kant's critical philosophy than is generally realized. However this may be, it is a matter of historical record that Whewell relied heavily on Kant's arguments for the *a priori* nature of space and time. History also records that Whewell stood almost alone in Britain (though Mansel's contribution cannot be overlooked) as the champion of Kant, trying hard to win a wider circle of friends in England and Scotland for the German philosopher's system.

Whewell's philosophical system, then, appears to be an imported hybrid, rather than a plant native to the soil in which it grew. But to come more

* A shorter version of this paper was read before the Cambridge University History and Philosophy of Science Seminar, Whipple Science Museum, February 25, 1963. I gratefully acknowledge the support given to my research by a National Science Foundation Postdoctoral Fellowship.

¹ In this paper, "empiricism" refers to that epistemology that holds that all non-trivial and non-analytic knowledge is based on experience, and that no such knowledge is certain or necessary.

² Dr. Walter Cannon of the Smithsonian Institution has pointed out to me in conversation that pre-Darwin Victorian English men of science—one thinks at once of Sir John Herschel and of Whewell in this context—held to an implicit "world picture" that included acceptance of an orderly, divinely designed universe, of physical science as the rational paradigm, and of mathematics as the language in which natural laws were originally written. Indeed, most of the key ingredients that went into seventeenth-century rationalism were to some extent reproduced in nineteenth-century England, especially in the thought of those associated with Cambridge. In one sense, then, Whewell's philosophy can be regarded as having made explicit what was implicit in the attitudes of some of his contemporaries.

quickly to the essential point: though nearly everything in his professional experience and in the tenor of his times ought to have led Whewell toward a militant philosophical empiricism, he spent his life defending a philosophy whose central contention was that inductive science yields universal and necessary truths. In a letter, Whewell stated that he regarded induction as his "special business"; in another letter, he calls induction "the true faith."³ Coming from one who was himself an inductive scientist, a great historian of inductive sciences, and a philosopher of science who introduced an articulate view of the nature of induction, these statements are perhaps understandable. What is not so easily understood is that, unlike many others (including Mill and Herschel among his contemporaries) who have had induction as their "special business," Whewell came to regard inductive inferences as demonstrative and as resulting in necessary truths. Indeed, Whewell's contention that what others regard as merely contingent empirical truths are in fact certainties was the hardest part of his philosophy for his contemporaries to accept; at the same time, it was the part of his philosophy that he took to be the most significant and unique.⁴ Up to and including the third edition of his *Philosophy of the Inductive Sciences*, that is, throughout his entire philosophical career, Whewell defended the proposition that science

develops in the direction of becoming a comprehensive system of laws that are both universal and necessary, and which are nevertheless in some sense the results of induction. Thus, to understand Whewell's theory of science, his views of necessity and induction must be seen clearly as fundamental within the system. In what follows, I shall attempt to throw into clear relief one of these views, namely Whewell's conception of necessary truth.

II. THE FUNDAMENTAL IDEAS

At first glance, Whewell's answer to the Kantian question, "How are *necessary* and *universal* truths possible?" is straightforward and simple.⁵ "... The necessity and universality of truths are derived from the *Fundamental Ideas* which they involve" (HSI, vol. I, p. 87).⁶ For Whewell, Fundamental Ideas "... are not Objects of Thought, but rather Laws of Thought. Ideas are not synonymous with Notions; they are Principles which give to our Notions whatever they contain of truth" (PIS, vol. I, p. 28; HSI, vol. I, p. 34). He also states that "... by the word Idea (or Fundamental Idea) used in a peculiar sense, I mean certain wide and general fields of intelligible relation, such as Space, Number, Cause, Likeness" (NOR, p. 187).⁷ Fundamental Ideas are what the activity of mind contributes to knowing. Whewell likens some of them, notably space, time,

³ To James Garth Marshall, Dec. 25, 1849; to Richard Jones, Aug. 21, 1834.

⁴ On Sept. 6, 1837, he wrote to Richard Jones asking him to "Put down on paper, as clearly and strongly as you can, the reasons which you can find for the opinion you held a little while ago; namely, that the simplest mechanical truths depend upon experience in a manner in which the simplest geometrical truths do not: that the axioms of geometry may be self-evident, and known *a priori*; but that there are not axioms of mechanics so known and so evident. I am very desirous of getting this opinion in its best and most definite shape, because the negation of it is a very leading point of my philosophy. This tenet separates me from the German schools as well as from the Scotch metaphysicians, and is the basis of a long series of results both speculative and practical. The whole *art* of induction depends upon it."

⁵ My account of Whewell's theory of the Ideas, and some parts of my account of his theory of necessary truth, parallel the account in C. J. Ducasse, "Whewell's Philosophy of Scientific Discovery I," *Philosophical Review*, vol. 60 (1951), pp. 56-59. This paper by Ducasse, together with his second paper on Whewell—"Whewell's Philosophy of Scientific Discovery II," *ibid.*, pp. 213-234 (both papers are reprinted as one under the title, "William Whewell's Philosophy of Scientific Discovery," in *Theories of Scientific Method: the Renaissance through the Nineteenth Century*, edited by R. M. Blake, C. J. Ducasse, and E. H. Madden [Seattle, 1960], pp. 183-217)—is the only reasonably accurate and detailed exposition of the fundamental elements in Whewell's theory of science written in this century. But Ducasse's purpose—to expose clearly the fundamentals of Whewell's philosophy of discovery—is more limited than mine. I wish to show both the essential features of Whewell's theory of necessary truth, and to say something about its development in the broader context of Whewell's philosophical thought. My account makes good one defect of Ducasse's work: it shows that Whewell did make some attempt to justify his otherwise purely psychologistic theory of necessary truth. Thus I hope to have gone beyond Ducasse's papers in the direction of a fuller exposition of both Whewell's theory and the context of discussion and thought in which it developed.

⁶ Capital letters are used throughout this paper to abbreviate the titles of Whewell's works. A complete list of Whewell's works cited appears at the end of the paper.

⁷ Whewell attempted to find synonyms for the term "idea" in a letter written to Richard Jones (Aug. 21, 1834): "I expect to shew clearly that in order to arrive at knowledge or science we must have, besides impressions of sense, certain mental bonds of connexion, ideal relations, combinatory modes of conception, scintial conditions, or whatever else you can help me to call them: they are what I called *Ideas* in my former letter. . . ." In his notebooks (1830-33) Whewell calls the Ideas "regulative conceptions," "interpretative conceptions," and "conditions of inductivity," *Induction*, I, II, IV. (Uncatalogued manuscript notebooks, Wren Library, Trinity College, Cambridge.)

and number, to Kant's *forms* of intuition. Others, for instance the ideas of cause and likeness, play for Whewell something akin to the role of Kant's categories, though he nowhere uses Kant's term so to designate them. Furthermore, in his treatment of some of the Fundamental Ideas, especially space and time, Whewell's account of their epistemological status deviates very little from the Kantian theory.⁸ Thus the Ideas of space and time *inform* our sensational experience (without being derivable from it), making meaningful perception possible. Whewell also speaks of Ideas as *subjective* forms for interpreting experience in such a way that knowledge-yielding statements about it become possible. Finally, the philosophy of each science consists in the development of the Fundamental Ideas that articulate and organize the propositions of that science, and give it whatever in the nature of truth status it might possess.

So regarded, Whewell's Fundamental Ideas are simply Kant's forms of intuition and categories under a new name. But what is significant for our purposes is not this evident similarity of his doctrine to Kant's (which Whewell readily admits), but rather the novel features of Whewell's position, which his subsequent discussion brings forth. For Whewell's central (and largely novel) contention is this:

The Progress of Science consists in a perpetual reduction of Facts to Ideas . . . Necessary Truths belong to the Subjective, Observed Facts, to the Objective side of our knowledge. Now in the progress of that exact speculative knowledge which we call Science, Facts which were at a previous period merely Observed Facts, come to be known as Necessary Truths; and the attempts at new advances in science generally introduce the representation of known truths of fact, as included in higher and wider truths, and therefore, so far, necessary. . . . Such steps in science are made, whenever empirical facts are discerned to be necessary laws; or, if I may be allowed to use a briefer expression, whenever *facts are idealized*. (FAII, pp. 33-35)

Now this view that empirically observed truths can become necessary ones, or as Whewell says

elsewhere, that *a posteriori* truths become *a priori* (PD, pp. 357-358) appears to be quite incompatible with the Kantianism of his general conception of the Fundamental Ideas. On the one hand Whewell wants to hold that there is a distinction between necessary and empirical truths, and thus that necessary truths cannot depend for their evidence upon appeal to experience; on the other hand, he wants to hold that necessary truths emerge *as necessary* in the course of the development of this or that empirical science. But if this latter is the case, it is difficult to see how necessary truths can be rigorously distinguished from empirical ones, and how they can be, as in the case of space and time, conditions residing in the constitution of the human mind to which all present and future experiences must conform.⁹

To understand this initially astonishing view, one must comprehend in detail both the nature of Whewell's Fundamental Ideas and the character of necessity that they bestow on some propositions of fact. It is clear that for Whewell there will be at least as many Ideas as there are sciences, and that each set of Ideas relative to a given science will make possible the expression of the laws of that science. But since sciences develop in concrete historical situations and over long periods of time, it follows that we do not now know every Idea that there is to be known. Thus Whewell's theory does not imply that the mind is pre-stocked with such Ideas and is therefore ready at once to develop particular sciences. Quite the contrary seems to Whewell to be the case.

It is not the *first*, but the most complete and developed condition of our conceptions which enables us to see what are axiomatic truths in each province of human speculation. Our fundamental ideas are necessary conditions of knowledge, universal forms of intuition; inherent types of mental development; they may even be termed, if any one chooses, results of connate intellectual tendencies; but we cannot term them *innate* ideas, without calling up a large array of false opinions. . . . Fundamental Ideas, as we view them, are not only not innate, in any usual or useful sense, but they are not necessarily *ultimate* elements of our

⁸ In HSI, vol. I, p. 87, Whewell states that his discussion of space and time contains "... the leading arguments respecting Space and Time, in Kant's *Kritik*." See also PD, p. 335.

⁹ Robert Blanché, in *Le Rationalisme de Whewell* (Paris, 1935), (p. 2), has argued that there are two fundamental problems in arriving at a fair interpretation of Whewell's philosophy. First, are ideas constitutive elements of the structure of reason, or contingent creations of individual genius? Second, are ideas forms in the Kantian sense, or substantive notions as in Descartes and Plato? Blanché thinks that Whewell never entirely answers either question, and thus that there is "... l'indécision fondamentale de sa philosophie, oscillant sans cesse entre l'idéalisme et le réalisme, entre l'épistémologie et l'ontologie." I hope to be able to show that Whewell does make a decision on both questions, though the attempted solution that results is hardly satisfactory.

knowledge. They are the results of our analysis so far as we have yet prosecuted it; but they may themselves subsequently be analysed. (DMH)

The position seems therefore to be something like this: man is subjected to a wide variety of sensations over many of which he exerts no form of control, and which constitute the "matter" of knowledge. But these sensations, in order to yield knowledge in the form of general propositions, must be *formed*, and each of the forms that we actively bestow upon experience becomes a candidate for scientific necessary truth. It is clear, however, that not every Idea will organize experience in such a way as to produce systematic and general knowledge that is expressed by means of natural laws. To get science, we need ideas that are clear and distinct, and that adequately colligate facts in such fashion that general propositions about matters of fact become possible. If, in turn, the general propositions (laws) support one another through what Whewell calls a "consilience of inductions,"¹⁰ and provide the basis for the deductive derivation of further truths, these general propositions (or the axioms from which they are derivable) will be seen intuitively to be not only true, but necessarily true, that is, their logical negations will be incapable of clear and distinct conception, "even by an effort of imagination, or in a supposition" (HSI, vol. I, p. 58).

III. THE NATURE OF NECESSARY TRUTH

Throughout his writings on the subject, Whewell insists upon two fundamental features of necessary truth. First, no necessary truth is derivable from experience, and second, a necessary truth is one whose negation is not only false, but impossible, that is, "necessary truths are those of which we cannot distinctly conceive the contrary" (FA, pp. 3-4). These two propositions characterizing necessary truth together imply a third: necessary truths are not known by means of discursive reasoning, nor are they, in Kant's sense of the term, analytic;

necessary truths are known by *intuition*. Each of these three propositions needs to be understood at some depth before Whewell's general position that facts are idealized (that truths of experience become necessary truths) can be made really clear.

When Whewell says that no necessary truth can be derived from experience, he seems at first glance to be saying something quite inconsistent with holding also that necessary truths emerge as necessary in the course of development of merely empirical sciences. The seeming inconsistency goes away, however, when we look at Whewell's arguments, instead of at his statements of the view. For the word "derived," which Whewell everywhere uses, does not fully express his intended meaning. What he does mean is that the *evidence* for necessary propositions is never empirical, that we cannot arrive at necessary truths by the simple expedient of collecting and listing facts. Thus, for example, we might be thought to be confirming a simple arithmetical proposition empirically by adding objects together to get a sum. But, thinks Whewell, as soon as we conceive the numbers—*not* the objects counted—distinctly, we *see* the sum, and no number of repeated trials of counting objects will alter the truth of the proposition thus seen. Indeed, "we cannot be said to make a trial, for we should not believe the apparent result of the trial if it were different" (FA, p. 16). Similarly, in discussing the statical principle, "the pressure on the support is equal to the sum of the bodies supported," Whewell makes the following point:

But in fact, not only are trials not necessary to the proof [of this proposition], but they do not strengthen it. Probably no one ever made a trial for the purpose of showing that the pressure upon the support is equal to the sum of the two weights. Certainly no person with clear mechanical conceptions ever wanted such a trial to convince him of the truth; or thought the truth clearer after the trial had been made. If to such a person, an experiment were shown which seemed to contradict the principle, his conclusion would be, not that the principle was doubtful, but that the apparatus was out of order We maintain, then, that this

¹⁰ NOR, pp. 70, 88-90. Aphorism XIV: "The *Consilience of Inductions* takes place when an Induction, obtained from one class of facts, coincides with an Induction, obtained from another different class. This consilience is a test of the truth of the Theory in which it occurs." Thus a consilience of inductions takes place when a hypothesis predicts facts of a kind different from those it was introduced to explain. In NOR, Whewell has made much of the fact that hypotheses (and hence the Ideas that they introduce) are *invented* by scientists, an admission that has led some to suppose that for Whewell hypotheses are *conventions*. This supposition is patently false, for Whewell insists that the conceptions introduced by hypotheses be clear and distinct, appropriate to the matter they are to deal with, and carefully verified by subsequent observations and experiments. Invention was for Whewell a most important step in *discovery*, but not in *proof*, which requires a patient and exact comparison of hypotheses with facts. Finally, Whewell thought that sound inductive proofs are demonstrative, so that some inductions show that the facts are expressible *only* by means of certain ideas and not by others. (NOR, pp. 111-112)

equality of mechanical action and reaction, is one of the principles which do not flow from, but regulate our experience. To this principle, the facts which we observe must conform; and we cannot help interpreting them in such a manner that they shall be exemplifications of the principle. (HSI, vol. I, pp. 217-218)

The result of this sort of argument is clear. No experience, even a highly organized and regularized experience brought about by an experiment, can confirm a necessary truth, because no conceivable experience could disconfirm it. Every apparent disconfirmation will either be traced to a mistake in procedure or will be interpreted to fit the law. Whewell here anticipates an important point that has received much attention in recent philosophy of science. No proposition can be regarded as empirically significant if there is no empirical evidence that could possibly disconfirm it. Such a proposition is either meaningless, or else analytic. However, though Whewell anticipates this point, he would not subscribe to it in its more recent form.

It seems then that Whewell's view that in the course of time empirically known propositions come to be apprehended as necessary cannot be interpreted to mean that the *logical* character of a proposition changes in time. Logically necessary propositions cannot be constructed out of logically contingent ones. (Which is not the same as saying that contingent propositions have no logically necessary implications.) Whewell's argument against such a construction is everywhere the same: contingent propositions that are learned from experience may be general, but they are never necessary. And since necessary propositions *are* necessary, it follows that they cannot be derived from experience.¹¹ Which means, of course, that the kind of evidence that establishes a necessary truth (for Whewell, intuition of its self-evidence), is different from the kind of evidence (observation or experiment) that establishes an empirical proposition. And this, for Whewell, is a position thought not to be incompatible with holding that truths known empirically come to be known subsequently as necessary.

Whewell was aware that the term "experience" is being used differently in the two sentences (1)

"Necessary truths are not derived from experience," and (2) "Some truths known by experience are later known to be necessary." In sentence (1) "experience" means for Whewell "observation or experiment in the context of some clear and precise scientific theory"; experience in this sense is thus regularized and deliberately accumulated. He speaks of this form of experience as the only one to which the name "can properly be applied," and says that "This experience is distinctive; it implies not only the faculty of perceiving, but special objects perceived; not merely the perception of something, but the perception that things are and occur in a certain manner to the exclusion of any other manner. It . . . [is] . . . contingent; it not only comes after the first exercise of perception, but it may fail to come at all . . ." (TD, p. vi). In sentence (2) "experience" means "sense experience" or simply "perception." It is in this second sense that one can say that necessary truths were known as matters of experience prior to being apprehended as necessary.

If it be said that we cannot possess the ideas of pressure and mechanical action without the use of our senses, and that this is experience; it is sufficient to reply that the same may be said of the ideas of relations in space; and that thus Geometry depends upon experience in this sense, no less than Mechanics. But the distinction of necessary and empirical truth does not refer to experience *in this sense*, as I need not now stop to show. (YM, p. vii)¹²

Though Whewell was convinced that there are these two separate and distinct meanings of the term "experience" he was not always careful to keep the two senses distinct in his writings. It may have been this lack of clarity in expression which confused some of his critics and gave rise to some of the difficulties in understanding his view of the apparent empirical origin of necessary truths. There are at least two places in his systematic writings, however, where Whewell does make the attempt to be as clear as possible on the two meanings of "experience." Whewell contends (NOR, pp. 59-63) that "Science begins with *Common Observation* of facts, in which we are not conscious of any peculiar discipline or habit of thought exercised in observing." He also refers to this form of observa-

¹¹ See, for example, FA; PIS, vol. I, pp. 62-66, 74-78; HSI, vol. I, pp. 65-68, 76-80.

¹² Cf. TD, pp. v-vii. It is this second sense of "experience" that Whewell employs in answering some of his critics. For example, he writes of Mill: "Mr. Mill cannot deny that our knowledge of geometrical axioms and the like, *seems* to be necessary. I cannot deny that our knowledge, axiomatic as well as other, *never* is acquired *without experience*." (IM, p. 79; PD, p. 286)

tion ("experience" in the second sense) as "... observation of the plainest and commonest appearances ..." which is the exercise of "... the mere faculties of perception." Experience in this sense is thus perception of appearances and recurrences of appearances of the most familiar things. It is in this sense of experience that we first observe the positions of the planets at different times, that we first become aware of the most obvious facts about bodies in motion, and that we first observe familiar aspects of visible objects. It is this sense of "experience" that Whewell thought to be in use in such sentences as "some truths known by experience are later known to be necessary," and "we learn by experience."¹³ Thus construed, experiences are perceptions of facts by individuals, and the nature of such facts will vary from individual to individual in that some perceptions will be clear, some confused, some will be of this aspect of a thing, some of another aspect of that thing. Given the at least partly idiosyncratic character of such perceptions they cannot by themselves count as evidence for any scientific generalization, nor for any proposition that is necessarily true.

If we are to have science, or what Whewell sometimes calls "speculative knowledge," it must be possible for us to have experience of the first type, namely controlled observation or experiment. In one place (PIS, p. 62), Whewell gives what is perhaps his clearest expression of what he means by "experience" in this sense. He writes:

I here employ the term Experience in a more definite and limited sense than that which it possesses in common usage; for I restrict it to matters belonging to the domain of science. In such cases, the knowledge which we acquire, by means of experience, is of a clear and precise nature; and the passions and feelings and interests, which make the lessons of experience in practical matters so difficult to read aright, no longer disturb and confuse us.

Whewell gives several examples of the kinds of propositions that we can know by means of such scientific experience. We know that animals which ruminate are cloven-hoofed, and that all the planets and their satellites revolve round the sun from west to east. Similarly we know by such intentional observation that all meteoric stones contain chrome. In order to have experience as scientific experience, Whewell believes that the

scientist must introduce clear and precise conceptions which, when they enter into the formulation of hypotheses about the actual course of events, allow us either to classify or to predict accurately. It was Whewell's view that the most efficacious scientific conceptions are those that permit us to make measurements or which permit deductive moves from one hypothesis to another in a scientific system. However, whether the conceptions introduced by the scientist are quantitative in character or not, they must at least permit the organization of data in such fashion that hypotheses which are either true or false of that data become possible. Thus what distinguishes such scientific experience from experience as mere perception is the fact that scientific experience, unlike mere perception, arises from the deliberate imposition of a concept on the data which when introduced into a hypothesis that is traced deductively to its consequences will permit the hypothesis to be either confirmed or refuted by precise observations or measurements which are not possible *until* the conception is introduced.

The distinction that Whewell seems to want is perhaps brought out by comparing the sentence (A) "I perceive *x* as red" and the sentence (B) "This *x* is red." Whewell would have to interpret sentence (A) as a report of what he calls a "mere perception." Given this interpretation, it would be true to say that "I learn by experience that a red appearance enters my visual field." In addition, I could perhaps characterize the experienced red as harsh or soft, hazy or clear, but I could not claim, at least on the basis of the experienced red alone, that the object seen is red. Sentence (B) is to be distinguished from sentence (A) in that sentence (B) is an objective knowledge claim which can be determined to be either true or false. In order to make such a determination, however, I would have to introduce a relatively clear and precise conception of color, such as is given in the conception of colors as wave-lengths of visible light. On the basis of this conception, I should then be able to determine whether or not the object seen has a color which falls within the wave-length range between 6220 to 7700 angstroms. Thus spectral analysis would make possible those "scientific experiences" that would determine whether the claim "this *x* is red" is true or not. The important point to bear in mind is that though the results of

¹³ See for example (PIS, p. 276) where, when speaking of necessary truths, Whewell writes "Truths *thus* necessarily acquired in the course of all experience, cannot be said to be learnt *from experience*, in the same sense in which particular facts at definite times, are learnt from experience, learnt by some persons and not by others, learnt with more or less of certainty."

precise scientific observation and measurement can provide evidence for scientific hypotheses, they can never completely establish these hypotheses as universal laws. Experiences as mere perceptions, on the other hand, though they may give rise to, or make us aware of states of affairs or propositions, never provide evidence for or against any objective knowledge claims. It is on the basis of this distinction that Whewell is able to hold both that necessary truths are learned by experience (in the sense of perception) and that these truths are not established as necessary by appeal to experience (in the sense of controlled scientific experience).

Whewell's two senses of "experience" and the resulting propositions to which they give rise may be regarded, in one sense, as interpretations of the meaning of "*Erfahrung*" in Kant's celebrated statement: "*Wenn aber gleich alle unsere Erkenntnis mit der Erfahrung anhebt, so entspringt sie darum doch nicht eben alle aus der Erfahrung*,"¹⁴ a statement with which Whewell's position surely agrees. In another and more vital sense, the two meanings of "experience" play fully operative roles in Whewell's theory. Given the meaning of "experience" as "scientific experience," Whewell was able to argue for his logical distinction between necessary and factual truths. Given the meaning of "experience" as "sensation" or "perception," he was able to hold that no knowledge is innate, and that thus even necessary truths take some time and experience to learn; and he was able to hold that, presupposing that no Ideas are ultimate, there are still more of them to be discovered in the future as new sciences become possible, and as subsequent discussion clarifies Ideas to such an extent that they become clear and distinct, and hence become sources of intuitively certain necessary propositions.¹⁵ It seems, then, that Whewell made out a perfectly good case for holding both that there is a logical distinction to be made between empirical and necessary propositions, and that some truths known empirically at first are subsequently known to be necessary.

In spite of his careful attempt to make a logical distinction between necessary and empirical statements, Whewell's position is somewhat puzzling in

that it is clear that he also wants to take statements asserting necessary truths to be non-trivial (non-analytic) and in some sense factually meaningful. Indeed, they are to be construed as the very source of meaningfulness in scientific systems, and they could not play this role, at least not so far as Whewell is concerned, unless they had some sort of reference to ontological realities. I think it obvious that Whewell does not want us to regard necessary truths as analytic, as true by virtue of being *logically* necessary. For his view that necessary truths are those whose negations are not clearly and distinctly conceivable is not translatable into the assertion that necessary truths are those from whose negations logical contradictions are derivable, even though this derivation might in fact be possible. In the course of his discussion with Mansel on the question of necessary truth, Whewell explicitly denies that the type of necessity characteristic of propositions expressing the Fundamental Ideas is logical necessity. "I will not pretend to say that this kind of necessity is exactly represented by any of those Fundamental Ideas which are the basis of science . . ." (PD, p. 342). Also to be taken into account is Whewell's continuing argument first against Stewart and then against Mill to the effect that both definitions and self-evident axioms are necessary as the basis of mathematical reasoning, and that hence mathematical reasoning is not hypothetical, but has an ontological reference. In addition, Whewell argues at some length that not even basic arithmetical and geometrical propositions are necessarily true in virtue of being logically necessary. Without actually referring to propositions in arithmetic and geometry as synthetic *a priori* propositions, Whewell does produce arguments that closely parallel Kant's arguments for the synthetic apriority of such propositions, including the arguments that the objects (space and time) of arithmetic and geometry are intuitive.

For example, Whewell argues that it is not true that the assertion "3 plus 2 equals 5" ". . . merely expresses what we mean by our words; that it is a matter of definition; that the proposition is an identical one" (a tautological one). Indeed, it is

¹⁴ Immanuel Kant, *Kritik der reinen Vernunft* [1781, 1787], B1.

¹⁵ None of Whewell's previous commentators have realized that his description of the avenues to the discovery (intuition) of necessary truth is throughout Platonic. We begin with confused and untrustworthy perceptions. These must first be rendered more orderly and less confusing by introducing precise scientific ideas. Discussion of the ideas plus accumulation of refined experiences eventually, so to speak, overcome the limitations of the "perceptual cave" in which we live, and we have then the power to intuit the necessity of some propositions whose ideas are clear and distinct. (Shades of the "Divided Line"?) Indeed, even Whewell's theory of induction has clear but frequently unrecognized Platonic motives.

not even true that the definition of 5 is "3 plus 2." Rather, the definition of 5 is "4 plus 1." But how is it that 3 plus 2 is the same number as 4 plus 1? "Not because the proposition is identical; for if that were the reason, all numerical propositions must be evident for the same reason. If it be a matter of definition that 3 and 2 make 5, it must be a matter of definition that 39 and 27 make 66. But who will say that the definition of 66 is 39 and 27? . . . How do we know that the product of 13 and 17 is 4 less than the product of 15 and 15? We see that it is so, if we perform certain operations by the rules of arithmetic; but how do we know the truth of the rules of arithmetic?" The correctness of the rules can be rigorously demonstrated. Perform this operation, and such-and-such must inevitably be the result.

Certainly this can be shown to be the case. And precisely because it *can* be shown that the result must be true, we have here an example of a necessary truth; and this truth, it appears, is not *therefore* necessary because it is evidently identical, however it may be possible to prove it by reducing it to evidently identical propositions. (HSI, vol. I, p. 59)

Thus Whewell does not want the second major characteristic of necessary truths—that we cannot distinctly conceive their negations—to be interpreted to mean that necessity is simply logical necessity. Rather, the necessity of propositions is to be traced back to the categorial necessity of the Fundamental Ideas that they express, and which form them. This brings us to the third essential feature of necessary truths, namely, that they are known by intuition of a certain kind. Whewell says,

. . . the way in which those Ideas became the foundation of Science is, that when they are clearly and distinctly entertained in the mind, they give rise to inevitable convictions or intuitions, which may be expressed as *Axioms*; and these Axioms are the foundations of Sciences respective of each Idea. (PD, pp. 336–337)

Whewell's critics misunderstood this view, which was not always clearly stated in many of his

expositions of it. He most especially does not mean that the Fundamental Ideas are innate possessions of every mind, nor does he mean that everyone in whatever condition of mental development will be able clearly and distinctly to apprehend necessary truths. Nor does the fact that some persons (children and mental defectives, for example) are unable to conceive clearly and distinctly the axioms of, say, geometry, mean that geometry thereby loses its universality and necessity. The attainment of a state of mind requisite for intuiting necessary truths is a cultural and educational development.¹⁶

Actually, Whewell's critics need not have been misled. Throughout his writings he insisted that the Ideas must be "clearly and distinctly possessed" before they can become sources of intuitively necessary axioms. He also tried to make it clear that the intuition of necessary truths is a "rare and difficult attainment" (PD, p. 339) coming only in that stage in the development of a science when the requisite categorial forms or Ideas shall have been simplified and organized (in short, rendered clear and distinct), *and* when properly trained and suitably ingenious scientists are available. And for Whewell this developmental thesis—"There are scientific truths which are seen by intuition, but this intuition is progressive" (PD, p. 344)—had a quite concrete historical meaning. No particular *result* of scientific investigation yields necessary truths. Rather, it is the *existence*, the *possibility*, of a science that establishes the necessity of its axiomatic principles (PD, pp. 349–350). Thus we know that particular Ideas *are* distinctly conceived when, in the course of actual history, they initiate inductive sciences that really do explain and predict the phenomena they were introduced to deal with. Those Ideas are really fundamental, which in fact organize and systematize whole bodies of general propositions; and such organized and systematized bodies of general propositions are precisely what for Whewell are to be counted as sciences.

More particularly, we know that a man has distinct conceptions if he can comprehend the axioms, and follow the reasoning, in any science either formal or empirical.¹⁷ Though this will mean that there are necessary truths that are not so

¹⁶ The developmental or progressive character of the intuition of necessity was an essential aspect of four of Whewell's theories: (1) his theory of necessary truth; (2) his theory of mathematical reasoning; (3) his theory of the historical development of science; and (4) his theory of education. For (1) see especially his answers to his critics in PD, chs. XXVIII, XXIX, For (2) see ME, appendix, "Remarks on Mathematical Reasoning and on the Logic of Induction." For (3) see HIS, HSI, and NOR. For (4) see TSM, pp. 5–33, and PEU, pp. 12–50.

¹⁷ This accounts for the central place Whewell accorded to training in mathematics in a liberal education. See TSM and PEU. Whewell's view of the intuitive necessity of the basic principles (Axioms) of the sciences also led him to attempt to construct statics, dynamics, and parts of other physical sciences as *deductive systems*. See TM and ME.

known to everyone, Whewell is not bothered by the seeming implication that certain people, namely those capable of clear conception, in effect create necessary truths which somehow depend for their necessity on these persons' ability to think clearly and distinctly. For when a man has knowledge, "We conceive that he knows it because it is true, not that it is true because he knows it. . . ." Further,

We are not surprized that attention and care and repeated thought should be requisite to the clear apprehension of truth. For such care and such repetition are requisite to the distinctness and clearness of our ideas: and yet the relations of these ideas, and their consequences, are not produced by the efforts of attention or repetition which we exert. They are in themselves something which we may discover, but cannot make or change. (FA, p. 27)

It follows from this that all Axioms are, in Kant's sense of the term, "constitutive" principles, or those principles without which knowledge would be impossible. Logic does not constrain us to think in the terms which the Axioms provide—we would not be guilty of simple logical contradiction if we entertained the concept of an uncaused event, for example—but our desire to have knowledge of reality does so constrain us. For surely we cannot *know* except by means of those principles which make knowing possible.

IV. NECESSARY TRUTH IN MECHANICS—AN ILLUSTRATION OF WHEWELL'S THEORY

For Whewell the paradigms of Ideas are the mathematical Ideas of space, time, and number, which are securely established as sources of the necessary axioms of geometry and arithmetic. Any additional Ideas that are to be admitted as the foundations of sciences must be distinctly conceivable in the same sense in which the mathematical Ideas are distinctly conceivable. The parallel between this doctrine and similar ones in Plato and

Descartes is obvious, though Whewell himself did not note it in the case of Descartes, whose system he misunderstood, perhaps as a result of his adulation of Newton.¹⁸ In any case, from the start of his philosophical career Whewell never doubted the central place of the mathematical Ideas in his developing theory of necessary truth.

It is also true that beginning early in his career Whewell took the Idea of equilibrium as expressed in the basic conceptions of statics as the source of the self-evidence of these conceptions.¹⁹ Indeed, the novel feature that Whewell introduced into his text in elementary mechanics was the sharp distinction between statics and dynamics. The distinction rested on two propositions. One, the laws of statics, unlike the laws of dynamics which involve proof by appeal to experiment, are simple and self-evidently true. Two, given this difference in the epistemological status of the two sciences, it is possible to develop statics in a completely deductive fashion, whereas in the case of dynamics the science requires development and establishment by appeal both to deductive proof and to experimental confirmation. Though no appeal to experiment is required in order to accept certain propositions in physics—for example, "The pressure on the support is equal to the *sum* of the masses supported" and "Fluids press equally in all directions"—an appeal to experiment is required in order to establish the truth of Newton's three laws. However, except for the purposes of teaching mechanics, Whewell was never quite satisfied with this distinction between statics and dynamics, and his dissatisfaction grew as the theory of necessity took firmer shape in the 1820's and 1830's. Even given the distinction, he did admit at an early stage that the laws of dynamics can be known *a priori* to be the most probable ones, and that they possess a certain simplicity and *seeming* self-evidence. He says in a note to the first edition of *Elementary Treatise on Mechanics* that

The undisputed authority which is now allowed to

¹⁸ For Whewell's interpretation of Plato's theory of ideas in terms of his (Whewell's) own conception of Ideas, see PTI, and PD, pp. 12 ff. For his almost incredible misunderstanding of Descartes, see HIS, vol. II, Bk. VII, pp. 140-145. Whewell was so intent on establishing Bacon and Newton as the heroes of the scientific revolution that he even criticized Descartes for holding a form of the view of science that Whewell himself held: "At the same time we may venture to say that a system of doctrine thus deduced from assumed principles by a long chain of reasoning, and not verified and confirmed at every step by detailed and exact facts, has hardly a chance of containing any truth. Descartes said that he should think it little to show how the world is constructed, if he could not also show that it *must* of necessity have been so constructed. The more modest philosophy which has survived the boastings of his school is content to receive all its knowledge of facts from experience, and never dreams of interposing its peremptory *must* when nature is ready to tell us what is." (Loc. cit.)

¹⁹ TM; see also ME, p. 159: "... I say that the axioms of Statics are *self-evidently true*."

the laws of motion . . . is the result of innumerable experiments never recorded and discussions now forgotten, to which they were subjected during the seventeenth century. Great numbers of trials were made, both by individuals and before learned bodies, to prove almost every one of the propositions which are now considered as nearly self-evident.²⁰

And from this admission that the axioms of dynamics are at least considered to be almost self-evident, he finally arrives at the position of the *Philosophy* that "... the whole science of Mechanics, including its most complex and remote results, exists as a body of solid and universal truths."²¹

This change in Whewell's view of dynamics appears to have been brought about by the realization that the laws of motion are based on the Fundamental Idea of *force*, a realization that was worked out in a communication to the Cambridge Philosophical Society in which Whewell discussed the nature of the truth of the three laws of dynamics, a paper which remained as his best discussion of the various roles played in science of the *a priori* and *a posteriori* elements in knowing and in coming to accept scientific laws.²² A detailed discussion of this paper might therefore serve both to reveal the details of this most promising of Whewell's attempts to analyze the logical structure of scientific laws and to illustrate his theory of necessary truth.

On Whewell's view, then, the Fundamental Idea upon which mechanics is based is the Idea of cause construed as force. The Idea of cause is partially expressed by three axioms of causality, which axioms are necessarily true, and, when the Idea of cause is "clearly apprehended," require no proof, and admit of none which makes them more evident (HSI, vol. I, p. 185). The axioms, in other words, are seen to be true by intuition by some scientists in the course of the history of mechanics. Whewell also states that such axioms as these are "not the result of any particular observations, but of the general observation or suggestion arising unavoidably from universal experience" (NTM, p. 9; TD, Preface, pp. v-vii).

The axioms expressing the meaning of causation are (1) "Every change is produced by a cause";

(2) "Causes are measured by their effects"; and (3) "Action is always accompanied by an equal and opposite reaction." Each of the axioms (which would have been called "synthetic *a priori* propositions" in another idiom) "... is necessarily true, and is a fundamental principle with regard to all mechanical relations." They also function as "governing and universal principles in all our reasoning concerning causes," and each "expresses a universal and constant conviction of the human mind." In short, "we inevitably and unconsciously assume the truth" of these axioms (NMT, pp. 2-4), and we make such an assumption because the three axioms of causality are each partial expressions of the meaning of the basic Idea of the cause-effect relation, which relation "... is a condition of our apprehending successive events, a part of the mind's constant and universal activity, a source of necessary truths; or, to sum all this in one phrase, a Fundamental Idea" (HSI, vol. I, p. 183).

The three axioms are expressions of the abstract Idea of Cause, which Whewell defines as follows: "By cause we mean some quality, power, or efficacy, by which a state of things produces a succeeding state" (HSI, vol. I, p. 173). The axioms, as expressions of this Idea of cause, will govern all possible thought in causal terms, no matter what the particular event caused and no matter what the particular cause. It is Whewell's view that we can in a sense instantiate these general axioms by supplying more particular conceptions of causes. Such instantiations or more precisely particularizations of the abstract axioms will then yield *a priori* laws of various types. For example, for the general idea of *any cause whatsoever* occurring in the axioms we can substitute the more particular or determinate conception of *cause of motion*, which Whewell defines as *force*. There remains the general conception of "cause as force" which in its turn will take on particular values or instances, though its generality is of a lower order than that of the idea of cause *per se*. Cause as force, says Whewell, is to be understood as "abstracted from all . . . special events, and considered as a quality or property by which one body affects the motion of the other" (HSI, vol. I, p. 184).

²⁰ TM, p. 264 note. The seeming self-evidence of dynamical laws is also admitted by Whewell in a number of early works, for example, in BT, pp. 231-232.

²¹ PIS, vol. I, p. 192; HSI, vol. I, p. 212. This might not seem much of a change in Whewell's position until one understands that for Whewell experience can render propositions *general*, but it cannot show them to be *universal*. Universal truths are necessary truths depending upon Fundamental Ideas. See PIS, vol. I, pp. 62-66; HSI, vol. I, pp. 65-67.

²² NTM. Much of the material of this essay is included in PIS, vol. I, pp. 177-185, 215-254, and in HSI, vol. I, pp. 184-204, 235-270. The essay was also reprinted as an appendix to the second edition of PIS.

Thus, when the causes we are considering are causes of motion (forces) we obtain by a form of instantiation three *a priori* laws of motion corresponding to the three causal axioms. These will be: (1') "When no force acts, the properties of the motion will be constant"; (2') "When a force acts, its quantity is measured by the effect produced"; and (3') "When one body acts upon another, there will be a reaction, equal and opposite to the action" (NTM, p. 6). Stated in this still somewhat abstract form, however, the laws, though of "absolute and universal truth . . . independent of any particular experiment or observation whatever," are "entirely useless and inapplicable" (NTM, p. 6). Whewell rightly thinks that this is so because in these forms the laws are simple logical derivatives of the axioms, following from these by simple and obvious reasoning. To be "useful and applicable" the laws, considered as formulae derived from *a priori* reasoning, must be expressed in such a way that experience can enter into their meaning; in other words, experience must be consulted for assignment of the values of the terms which enter into the formulae (NTM, p. 26). Throughout his writings on this subject, Whewell insists that the primary role of experience (in the special sense of that sort of experience that experiment and scientific observation makes possible) is to help us to illustrate and to understand these *a priori* laws, and others in different sciences.

The course of real knowledge is, to obtain from thought and experience the right interpretation of our general terms, the real import of our maxims, the true generalizations which our abstractions involve. (HSI, vol. I, p. 268)

Whewell next proposes to state the three *a priori* laws as causal laws forming the basis of the science of mechanics. The laws will now read:

(1") A body not acted upon by any force will go on in a straight line with an invariable velocity; (2") When a force acts upon a body in motion, the effect is the same as that which the same force produces upon a body at rest; and (3") In the direct mutual action of bodies, the momentum gained and lost in any time are equal. (NTM, pp. 7-15)

Now the truth of these laws (the familiar Newtonian laws) follows from the truth of the *a priori* laws of which they are further particularizations (and hence indirectly from the truth of the three

causal axioms), but *only if* we are entitled to suppose that all changes of motion result from *external* causes, that is to say, only if we can show by appeal to experience that changes in the motion of a body are not brought about by circumstances intrinsic to that motion itself. It is clear then that each of the empirical laws of motion will have an *a priori* form and an experiential content. Otherwise, that is in the absence of empirical content, each law of motion will remain undecidable so far as its claim to telling the truth about the world is concerned. It is thus in order to determine the truth or falsity of the supposition that all changes in motion are caused by forces external to the body in motion that we must have recourse to experiment. In the case of law (1"), results of experiment must be appealed to in order to determine whether or not the regularity stated by this law is time-dependent. For if the body is not acted upon by a force, no *spatial* relations cause its position. There could be no cause except one depending upon *time*, such as would be the case if bodies had a natural tendency to move slower and slower, according to a rate depending on the time elapsed. Experiments conducted with reference to the first law of motion show that the time for which a body has already been in motion is not a cause of change in the velocity of that body, and hence the law is confirmed (NTM, p. 7).

It follows from this analysis that for Whewell each of the three laws of motion must have both an *a priori* and an empirical part, the empirical part in each case amounting to the negation of the supposition that the cause of motion of a body is conditioned by changes in its motion, a negation which shows that the motion of a body is wholly dependent upon forces extraneous to the body itself (NTM, p. 22). The necessary component of the first law (1") is "Velocity does not change without a cause" (which follows logically directly from (1')), and the empirical component is expressed by the statement: "The time for which a body has already been in motion is not a cause of change of velocity." In the case of law (2"), the *a priori* part is "the accelerating quantity of a force is measured by the acceleration produced" (a logical consequent of (2')), and the empirical part is "The velocity and direction of the motion which a body already possesses are not, either of them, causes which change the acceleration produced," a proposition established by experiment. Finally, the necessary part of law (3") is "Reaction is equal and opposite to action" (logically derivable from (3')),

and the part established by experiment is "The connection of the parts of a body, or of a system of bodies, and the action to which the body or system is already subject, are not, either of them, causes which change the effects of any additional action" (NTM, p. 22).

For Whewell, then, appeal to experience gives meaning to the otherwise purely *a priori* laws of motion, making them at last "useful and applicable," and producing "real knowledge." One way of construing Whewell's analysis is to take him to mean that the empirical aspects of the laws contribute the *semantics* or the *interpretation* of these *a priori* forms, where the terms "semantics" and "interpretation" are used in their current philosophical significations. His analysis of the laws of motion does strongly suggest that he had in mind a primitive form of the idea that analysis of the logical structure of a science reveals that it contains both a calculus (the *a priori* logical forms of the statements involved plus rules for their logical manipulation) and an empirical interpretation of that calculus. But thus to construe Whewell's view is to leave the door open for serious misunderstandings. For, as has already been stated, Whewell does not interpret "necessity" in terms of "logical necessity," or "analyticity," and the *a priori* forms of the laws of motion must be regarded as necessary truths precisely in Whewell's sense of "necessity." Indeed, what interests Whewell is not straightforward analysis of the logical structure of scientific laws, but rather the epistemology and the metaphysics of scientific systems. He wants to know why science results in real (that is, certain) knowledge, and he also wants to know why it is that the necessary forms of thought do actually apply to systematic scientific experience.²³ And what he says on this matter with particular reference to the laws of motion is odd and at first glance perplexing.

For Whewell's view is that, in general, what we learn from the results of experiments performed to decide the truth of the three laws of motion is that

Instead of having to take into account all the circum-

stances of moving bodies, we find that we have only to reject all these circumstances. Instead of having to combine empirical with necessary laws, we learn empirically that the necessary laws are entirely sufficient. (NTM, p. 23)

Thus, though the appeal to experience does not contribute evidence that strengthens the necessary truth of the laws (no empirical evidence could do *that*), the appeal does show that it was itself superfluous. Experience does not establish the necessary truth of the laws (only the axioms can do this), but it does illustrate that what we saw intuitively to be the case is in fact the case. And for Whewell this means, at least in the case of well-articulated sciences like mechanics, that we can now develop the science in question as a *deductive* system, experimentation having shown itself to be irrelevant. One might even claim with justice that the central point of Whewell's view of inductive science is that experimentation only exists for the sake of the eventual abolition of laboratories. Put somewhat less colorfully, Whewell's point is that the laws, once experimentally accepted, appear simple and self-evident, and hence appear not to be in need of further experimental confirmation. For this reason we can take experiments as playing the role of rendering the laws more clear and intelligible, "... as visible diagrams in geometry serve to illustrate geometrical truths" (HSI, vol. I, pp. 248-249). For Whewell, the important conclusion of this analysis is the following one:

And thus we see how well the form which science ultimately assumes is adapted to simplify knowledge. The definitions which are adopted, and the terms which become current in precise senses, produce a complete harmony between the matter and the form of our knowledge; so that truths which were at first unexpected and recondite, become familiar phrases, and after a few generations sound, even to common ears, like identical propositions . . . [All sciences illustrate] . . . the general transformation of our views from vague to definite, from complex to simple, from unexpected discoveries to self-evident truths, from

²³ The basic problem of Whewell's philosophy was for him the establishment of the truth of the proposition, "Man is the Interpreter of Nature, and Science is the right Interpretation" (PIS, vol. I, p. 37; HSI, vol. I, p. 41). Thus it might be said that he was seeking adequate "bridge rules," "epistemic correlations," "semantical rules," or "rules of interpretation" that would provide an epistemological link between theoretical scientific terms and experiments, or between theoretical laws and empirical laws. But given his theory that there are (synthetic *a priori*) axioms that are the ultimate epistemological justification-sources of the sciences, his solution to the bridge rule problem (the problem of showing that theoretical systems say something true about the empirical world even though theoretical terms have no direct empirical referents) takes, in the end, a partly Kantian, partly theological form, as will be shown later in this paper.

seeming contradictions to identical propositions. . . . (HSI, vol. I, pp. 260-261)²⁴

Whewell's position on the laws of motion may perhaps be best summarized in his own words:

What is the source of the axiomatic character which the propositions [laws of motion] thus assume? . . . The laws of motion borrow their axiomatic character from their being merely *interpretations* of the Axioms of Causation. Those axioms, being exhibitions of the Idea of Cause under various aspects, are of the most rigorous universality and necessity. And so far as the laws of motion are exemplifications of those axioms, these laws must be no less universal and necessary. . . . What should happen universally, experience might be needed to show: but that what happened should happen *universally*, was implied in the nature of knowledge. The universality of the laws of motion was not gathered from experience, however much the laws themselves might be so. . . . The laws of motion borrow their *form* from the Idea of Causation, though their *matter* may be given by experience: and hence they possess a universality which experience cannot give. They are certainly and universally valid; and the only question for observation to decide is, how they are to be understood. (HSI, vol. I, pp. 265-267)

Furthermore, experience is able correctly to interpret these ideational forms, precisely because they are *conditions* of experience which are exemplified in particular cases. "Experience is the interpreter of nature; it being understood that she is to make her

interpretation in that comprehensive phraseology, which is the genuine language of science" (HSI, vol. I, p. 268). The answer is straightforward in its evident Kantianism. Experience gives us knowledge of causal connections, because knowledge-yielding experience is only possible as causally connected. If the only kinds of questions that we can ask nature are causal questions, it ought not to seem surprising that the only answers we get are causal answers. The same sort of thing is true, of course, of other Fundamental Ideas and their connection with experience.

V. THE JUSTIFICATION OF THE IDEAS— WHEWELL'S "ULTIMATE PROBLEM"

Whewell's first attempt to justify the necessity of some of our scientific laws thus takes a fairly clear Kantian form. Experience confirms laws of certain forms only because experience cannot take place in any other forms. And at least in so far as the Ideas of space, time, number, and cause are concerned, Whewell apparently wished to rest content with this Kantian answer, at least up to a certain point. But what caused both Whewell and his critics difficulties was the fact that he wanted to extend the number of *a priori* forms to include Ideas in the less well-developed sciences of chemistry and biology, and even in geology and natural history. Indeed, Whewell had enough philosophical humility even to want to leave the question open whether or not we would in future discover other,

²⁴ Whewell was aware that his position that "... all that we learn from experience is, that she has nothing to teach us concerning the laws of motion . . ." was similar to d'Alembert's position. (NTM, p. 24) Surely Whewell could have agreed with d'Alembert's statement of the problem of necessity in science in *Traité de dynamique* (1758), p. xxiv: "The great metaphysical problem has been put recently: are the laws of nature necessary or contingent? To settle our ideas on this question, we must first reduce it to the only reasonable meaning it can possibly have . . . viz., whether the laws of equilibrium and motion that we observe in nature are different from those that matter would have followed, if abandoned to itself. Hence this is the way the scientist should follow: first he should try to discover through reason alone which would be the laws of mechanics in matter abandoned to itself; then he should investigate empirically what are really such laws in the universe. If the two sets of laws be different, he should conclude that the laws of mechanics, such as those yielded by experiment, are of contingent truth, since they appear to spring from a particular and express decision of the Supreme Being: if, on the other side, the laws yielded by experiment agree with those that could be deduced by logic alone [the position of Whewell, and, in a somewhat different form, of d'Alembert], he shall conclude that those laws are of necessary truth: which does not mean that the Creator could not have established a wholly different set of laws, but that he did not hold it right to establish other laws than those which resulted from the very existence of matter." Quoted in Georgio de Santillana, "Aspects of Scientific Rationalism in the Nineteenth Century," in *International Encyclopedia of Unified Science*, vol. II, no. 8 (Chicago, 1950), p. 10. For an interpretation of d'Alembert's rationalistic philosophy of science see Robert E. Butts, "Rationalism in Modern Science: d'Alembert and the *Esprit Simplicité*," *Bucknell Review*, vol. 8 (1959). As will emerge shortly from the exposition of this paper, even the theological framework of d'Alembert's posing of the question of necessary truth can be seen to be of vital importance to Whewell's later position on the justification of necessary truths.

at present unknown, necessary truths.²⁵ Some critics have expressed surprise that Whewell was not perceptive enough to see that his Fundamental Ideas required something like a "transcendental deduction" to justify them as the conditions of all reliable knowledge.²⁶ But such surprise betrays a fundamental misunderstanding of Whewell's intentions and positive doctrines. The intuition of necessary truth is progressive; we cannot be sure, then, that we now know all of the Fundamental Ideas and their associated axioms that will perhaps later be seen to be the foundations of sciences. How, under circumstances of this sort, can we provide a dogmatic (or even a "critical") justification of the conditions of knowledge? We simply have to wait to see what indeed will be the necessary conditions of future theoretical systems.

Thus it can be no mystery that Whewell, holding as he did that the intuition of necessary truth, the idealization of facts, is progressive, did not provide a transcendental deduction of the Ideas. But this does not mean that his mind was free from philosophical anxiety about the question of the final justification of those propositions whose necessity has been intuited. Indeed, he tells us of his concern in a letter to Herschel.

Ideas and Things are constantly opposed, yet necessarily co-existent. How they are thus opposite and yet identical, is the ultimate problem of all philosophy. . . . Knowledge requires Ideas. Reality requires Things. Ideas and Things co-exist. Truth is, and is known. But the complete explanation of these points appears to be beyond our reach.²⁷

In part, Whewell thought that the solution of this ultimate problem of all philosophy is beyond our

reach because he would not settle for a merely epistemological justification of the Fundamental Ideas. In short, Whewell's Kantianism stops being operative at exactly the point where he thinks that only a metaphysical account will give an answer to this ultimate problem of philosophy.

Whewell had maintained that the chief result of his philosophy of science was the establishment of the claim that science is not only one among many alternative systems for conceptualizing reality, but that it is in addition the *only* correct conceptual system. "Man is the Interpreter of Nature, Science the right interpretation" (NOR, p. 5). This position surely implies one form of scientific realism, and such a metaphysics, if it is to be held at all, must be argued for on grounds other than those provided by particular results of the sciences. But Whewell's philosophy of science had emphasized the role of the human mind as an active agent imposing ideas upon sensuous materials. This stress on the processes of rational construction and system building involved in the work of scientists convinced some of Whewell's contemporaries that his philosophy of science was underwritten either by an uncritical acceptance of Kant's idealism, or by a mere conventionalism of the Ideas. Thus for Whewell the epistemological justification of the special status of the Ideas in human knowledge had to take the form of a more basic justification of his philosophy of science as realistic.

In the letter to Herschel mentioned above, Whewell quotes from a review of his *History and Philosophy* that appeared in *Quarterly Review* in June, 1841. The reviewer had taken Whewell to task for suggesting that space is merely an idea, a proposition which for the reviewer implied that

²⁵ In his notebooks (uncatalogued mss. in Wren Library, Trinity College), Whewell lists the Ideas that will later play a key role in the *Philosophy*: space, time, number, cause, opposition, resemblance, substance, etc. (*Induction*, I, II). In *Induction* I he also lists "regulative conceptions" of political economy—property, exchangeable value, labor, etc., and in *Induction*, IV, under the date Apr. 12, 1831, he lists the "regulative or interpretative conceptions" of "notional" (subjective) sciences—character, countenance, wisdom, God, right, pleasure, etc., which are Ideas that come to play a large role in Whewell's theory of morality (see EM, vol. I, p. 50; vol. II, p. 6). Whewell's main contention is that each science must be based upon such Ideas, but he was perfectly willing to admit that what he suggested as ideal bases of some sciences, namely the imperfectly developed ones, might not turn out to be distinctly intuitable in the course of history. A recent commentator (Harold T. Walsh, "Whewell on Necessity," *Philosophy of Science*, vol. 29 [1962]) has been misled by this tentativeness of Whewell's attitude into thinking that Whewell took the Ideas to be sources of merely hypothetical or relational necessity (pp. 141, 145) a position Whewell explicitly denies in PIS, vol. I, p. 100; HSI, vol. I, p. 101. Mill also misunderstood the tentative nature of the attitude that Whewell took toward some of the Ideas, like the Ideas of definite composition and definite quantity in chemistry. Whewell explained that ". . . what I meant to do [in the *Philosophy*] was, to throw out an opinion, that if we could conceive the composition of bodies *distinctly*, we might be able to see that it is necessary that the modes of composition should be definite. . . . The thought of such a necessity was rather an anticipation of what the intuitions of philosophical chemists in another generation would be, than an assertion of what they now are or ought to be. . . ." (PD, p. 340). For Whewell mathematical physics was the paradigm science, and he was far from thinking that other sciences matched exactly its intuitive certainty.

²⁶ Walsh, *op. cit.*, p. 141; C. J. Ducasse, *op. cit.*

²⁷ Letter to Sir John Herschel, Apr. 11, 1844 (uncatalogued mss. copy, Wren Library, Trinity College, Cambridge). Printed as an appendix in PIS and PD.

things do not actually exist in space and therefore that space is not a reality, but "... a mere matter of convention or imagination." Whewell was quick to combat this reviewer's opinion by arguing that his theory of the Ideas did not commit him to the view that the external world is not real. In arguing against the position of the reviewer in the letter to Herschel, Whewell first states his view of the inseparability of facts and Ideas. "Our *real knowledge* is *knowledge*, because it involves Ideas, *real*, because it involves Facts. We apprehend things as existing in space because they do so exist: and our idea of space enables us so to observe them, and so to conceive them." However, in the same letter Whewell admits that the reality of perceived objects is "... a profound, apparently an insoluble problem." He regards it as impossible for us to suppose that there is not something real independently of our knowledge, but recognizes also the peculiarity and egocentricity of the knowing situation. "Yet how can we conceive truth otherwise than as something known? How can we conceive things as existing without conceiving them as objects of perception?" It is at this point that Whewell introduces what he calls "the ultimate problem of all philosophy," the statement of which is quoted above.

Thus it seems clear that Whewell was bothered by the realization that his philosophy of science had not been given the kind of metaphysical justification that he himself recognized as required. Furthermore the problem at the basis of his system that he recognizes in the letter to Herschel began to disturb him as early as 1841, that is, one year after the appearance of the first edition of the *Philosophy*. According to a footnote appearing in the published form of the letter to Herschel referred to above, Whewell's remarks about the difficulties involved in solving the problem of the reality of perceived objects were written in 1841, but Whewell did not send this letter to Herschel until 1844, when he sent with it a copy of his first memoir on "The Fundamental Antithesis of Philosophy" (PIS, p. 676). Three things are clear then: one, that Whewell was aware of the justification problem as related to his own system; and two, that he had thought about this problem as early as 1841; and three, that the first *Memoir* represented for Whewell an initial attempt to deal with this problem.

Without taking into account Whewell's letter to Herschel one cannot fully appreciate that the two memoirs on the fundamental antithesis of philosophy were for Whewell a preliminary attempt to solve what he called "the ultimate problem of all philosophy."²⁸ These memoirs²⁹ must be understood as embodying Whewell's argument that we can retain his doctrine of the Ideas as sources of necessary truth in science without committing ourselves either to Kantian idealism or to conventionalism; that is, that we can retain a realistic metaphysical basis for our philosophy of science by identifying in some way facts and ideas, perceptions and concepts. That Whewell took this doctrine seriously as required to sustain and improve his earlier philosophy of science is further evidenced by the fact that the first memoir "On the Fundamental Antithesis of Philosophy" was printed as an appendix in both the second and third editions of the *Philosophy* and by the fact that Whewell completely revised Book I of the *Philosophy* in the second edition of 1847, so that his position on the fundamental antithesis could be more fully incorporated. Indeed, in the preface to the second edition of the *Philosophy*, Whewell wrote

I have made very slight alterations in the first edition, except that the First Book is remodelled with the view of bringing out more clearly the basis of the work; this doctrine of the Fundamental Antithesis of Philosophy. This doctrine, and its relation to the rest of the work, have become more clear in the years which have elapsed since the first edition. (PIS, p. xi)

Turning then to Whewell's two memoirs on the fundamental antithesis of philosophy, regarded as papers addressed to the idealism-realism and the conventionalism-realism tensions, we find Whewell discussing the traditional philosophical propensity to systematize bifurcations of both epistemological and ontological sorts. The history of philosophy exhibits many systems that distinguish necessary from experiential truth, facts from theories, ideas from sensations, thoughts from things, deduction from induction, and so on. And though, at least for the purposes of candid theory of knowledge, Whewell finds such distinctions helpful, he argues in the memoirs that no such distinction, no such fundamental antithesis, can be held to be ultimate. Thus he writes:

²⁸ For example, Walsh, *op. cit.*, pp. 141-142, completely misrepresents the role that the two memoirs on the fundamental antithesis of philosophy play in the development of Whewell's philosophy. I hope in what follows that I can put this matter right.

²⁹ FA and FAIL.

But though philosophy considers these elements of knowledge separately, they cannot really be separated, any more than can matter and form. . . . No apprehension of things is purely ideal: no experience of external things is purely sensational. If they be conceived as *things*, the mind must have been awake to the conviction of things by sensation: if they be *conceived* as things, the expressions of the senses must have been bound together by conceptions. If we *think* of any *thing*, we must recognize the existence both of thoughts and of things. *The fundamental antithesis of philosophy is an antithesis of inseparable elements. . . . The terms which denote the fundamental antithesis of philosophy cannot be applied absolutely and exclusively in any case. . . .* The Facts are Facts so far as the Ideas have been combined with the sensations and absorbed in them: the Theories are Theories so far as the Ideas are kept distinct from the sensations, and so far as it is considered as still a question whether they can be made to agree with them. A true Theory is a Fact, a Fact a familiar Theory. (FA, pp. 9-10)

It follows from this identification of the terms of the antithesis (for example, "necessary truth" and "empirical truth") that in the end such terms must be in some way derivable from one another (FA, p. 14). There can be no empirical scientific truth that is not conditioned by necessary axioms expressing Fundamental Ideas, and there can be no intuition of necessity without an experience of the things that will interpret and render understandable the necessary axioms.

I think it must be admitted that such an attempted identification of the terms of the fundamental antithesis takes us far beyond the limits of Kant's critical philosophy. For if we are to take Whewell seriously, not only must each of his Fundamental Ideas be taken as a constitutive principle of knowledge, it must also be taken as a non-subjective, extra-mental ontological reality. Thus, for Whewell, the solution of the ultimate problem of philosophy must be, and precisely in Kant's

sense of the term, *metaphysical*.³⁰ The Kantian transcendental deduction, the outlines of which Whewell accepted in the case of space, time, number, and cause, must now give way to what for him will finally be a *theological* deduction of all the Ideas. He hints at this in the second memoir on the fundamental antithesis when he suggests that it is man's special place in the creation that finally guarantees his knowledge.³¹ And in two later works, *Philosophy of Discovery* and *Of the Plurality of Worlds*, Whewell finally states the important "special results" of his theory of progressive intuition and his theory of Ideas.

In *Philosophy of Discovery* (pp. 354-358) Whewell at last formulates his central problem in detail.

How can there be necessary truths concerning the actual universe? . . . *How* can facts be idealized? How can that which is a fact of external observation become a result of internal thought? How can that which was known *a posteriori* become known *a priori*? How can the world of Things be identified with the world of Thoughts? How can we discover a necessary connexion among mere phenomena? . . . To put it otherwise: How is it that the deductions of the intellect are verified in the world of sense? How is it that the truths of science obtained *a priori* are exemplified in the general rules of facts observed *a posteriori*? How is it that facts, in science, always do correspond to our ideas? . . . It being established, then, that in the progress of science, facts are idealized—that a *posteriori* truths become *a priori* truths; that the world of Things is identified with the world of Thoughts to a certain extent; to an extent which grows larger as we see into the world of Things more clearly; the question recurs which I have already asked: How can this be?

Notice how crucial is the fact that Whewell even *recognized* this problem. In his youth he had proclaimed that the inductive philosophy is the "true faith," only to discover in his mature years that however true the faith is, however astonishing its

³⁰ I do not think that Whewell was explicitly aware how far his final attempted justification of the Ideas would depart from his earlier Kantianism. Part of the puzzle rests on the fact that Whewell's own definition of "metaphysical" underwent several significant changes. In the early 1820's (e.g., letter to Richard Jones, Sept. 23, 1822), he wanted to be rid of the term "metaphysical" because what it signified stood in opposition to the inductive philosophy. He also condemned seventeenth-century metaphysicians for using what Bacon had called the "method of anticipation," a condemnation with which Kant's philosophy agrees. But during the same period he wrote the following in a letter to Jones (Aug. 16, 1822): "And so by the *metaphysics of mathematics* I mean the examination of the laws and powers of the mind on which their evidence depends, the analysis of their principles into the most simple form and if you choose the history of their development. It is not easy to stick to the distinction between this and the *logic* of the science; but the latter examines the accuracy of your mode of deducing consequences from your principles and the former your way of getting your principles." In NOR (p. viii) he wrote: "Metaphysics is the process of ascertaining that thought is consistent with itself; and if it be not so, our supposed knowledge is not knowledge." But Whewell never fully recognized anything like Kant's distinction between critical theory of knowledge and metaphysics as empirically empty speculation, perhaps because as his thought developed he saw more clearly that the consistency of thought with itself could only be established on completely non-Kantian grounds.

³¹ FAII, "Continuation of the Second Memoir." Originally printed for private circulation.

results are, the faith nevertheless requires justification. For the method of induction and the philosophy to which it gives rise are neither of them *self-critical*. They must be supplemented by, paradoxically enough, a return to exactly that form of philosophy—classical modern metaphysics—that had been regarded as the arch-enemy of the “true faith.”

So Whewell now tells us that the epistemological problem that he so fully expressed in *Philosophy of Discovery* can be answered if we can answer a more basic question, namely, the question: How did the world come to be what it is? (PD, p. 358). The supposition seems to be that if we can know what the world is, we can also know how it is possible for us to know anything at all. It is at this point, surely, that Whewell the Kantian becomes Whewell the seventeenth-century metaphysician. In fact, however, the answer to Whewell's metaphysical question was ready-made. For Whewell never doubted that the world was the world as the Christian theist understood it. The world came to be what it is through the creative agency of a supremely intelligent, supremely good, and supremely concerned Deity. Everything that follows from this position finally falls neatly into place and saves Whewell's faith in induction. What Kant and the early Whewell taught us to regard as constitutive epistemological conditions are now to be seen as constitutive *ontological* ones. Man as an intellectual creature can speculate about the relations of things to one another. In their most general forms, these relations will be such relations as space, time, and causality. Man “. . . can discover truths, to which all things, existing in space and time, must conform. These are conditions of existence to which the creation conforms, that is, to which the Creator conforms; and man, capable of seeing that such conditions are true and necessary, is capable, so far, of understanding some of the conditions of the Creator's workmanship. In this way, the mind of man has some community with the mind of God. . . .” (PW, p. 109)

This contention that man is capable of glimpses

of the Divine ideas is not unlike Galileo's view that man's understanding, though it *extends* to very few real truths, nevertheless has some truths in an *intensively* complete and perfect way; and in the latter case man's understanding approximates completely Divine understanding.³² Something akin to Galileo's distinction between extensive and intensive understanding is involved in the following passage in the essay *Plurality of Worlds*, in which Whewell summarizes his solution of the problem that had plagued him.

The Universe was created by its Author, in space; that is . . . space was one of the Fundamental Ideas involved in its construction, so that all its parts have relations which flow from this Idea, and cannot exist without having such relations; and . . . further, the mind of man partakes of this Idea of Space, and apprehends all objects according to this Idea, and cannot apprehend them otherwise. As in the Creative Mind, all things are regarded according to the Universal Creative Idea of Space; so in the created mind, having this Idea as part of its nature, all things are regarded according to the Universal Regulative Idea of Space. To man, the relations of space are seen as necessary relations, because they are a portion of the supreme and original act by which the Universe was made what it is; and because also the mind of man is made a sharer in the conditions of the creative act. To man, space is found to be a necessary condition of the existence of objects, because God has created all objects under this condition, and has made man capable of apprehending the universality of the condition. Man can apprehend no object which is not in space, because God has created no object which is not in space: the act of Creation involves a Universal Thought of the Divine Mind, which thought, the mind of man can, according to its human powers, admit and entertain. . . . The work of creation is, thus, subjected to conditions; and some of these conditions man apprehends as necessary truths. How many classes of such necessary truths there are, or can be, we need not here inquire. . . . Yet of some of these Ideas he can obtain glimpses; and these glimpses appear, in some cases at least, under the aspect of necessary truths; truths which not only are true, but, by the nature of things, *must be* true. And so far, even as seen

³² Galileo Galilei, *Dialogue on the Great World Systems*, Salisbury translation, revised by Georgio de Santillana (Chicago, 1953), p. 114: “. . . We should have recourse to a philosophical distinction and say that the understanding is to be taken two ways, that is *intensively* or *extensively*. *Extensively*, that is, as to the multitude of intelligibles, which are infinite, the understanding of man is as nothing, though he should understand a thousand propositions; for a thousand in respect of infinity is but as zero. But as for the understanding *intensively*, inasmuch as that term imports perfectly some propositions, I say that human wisdom understands some propositions as perfectly and is as absolutely certain thereof, as Nature herself; and such are the pure mathematical sciences, to wit, Geometry and Arithmetic. In these Divine Wisdom knows infinitely more propositions, because it knows them all; but I believe that the knowledge of those few comprehended by human understanding equals the Divine, as to objective certainty, for it arrives to comprehend the necessity of it, than which there can be no greater certainty.”

by man, contemplative truth and fact, possibility and actuality, present themselves as identical.³³

Whewell's major problem had been the one of justifying the apparently astonishing position that the intuition of necessity, the idealization of fact, was progressive, and that there are, nevertheless, necessary truths that are the epistemological foundations of the several sciences. And this problem was absorbed into the problem alluded to in the letter to Herschel, the problem of explaining that our intuited ideal forms are not empty, but really do catch reality, really do operate in the perception of real objects. In fact, these were for Whewell two forms of the same question, as the passage from *Philosophy of Discovery* (pp. 354-358) quoted above indicates. In the letter to Herschel Whewell hesitated and said only that a complete explanation is not now within our reach. In the following summary of his doctrine of the fundamental antithesis of philosophy he also hesitated:

In every act of knowledge (1) there are two opposite elements which we may call Ideas and Perceptions; but of which the opposition appears in various other antitheses; as Thoughts and Things, Theories and Facts, Necessary Truths and Experiential Truths; and the like: (2) . . . our knowledge derives from the former of these elements, namely our Ideas, its form and character as knowledge, our Ideas of space and time being the necessary forms, for instance, of our geometrical and arithmetical knowledge; (3) and in like manner, all our other knowledge involving a development of the ideal conditions of knowledge existing in our minds; (4) but . . . though ideas and perceptions are thus separate elements in our philosophy, they cannot, in fact, be distinguished and separated, but are different aspects of the same

thing; (5) . . . the only way in which we can approach to truth is by gradually and successively, in one instance after another, advancing from the perception to the idea; from the fact to the theory; from the apprehension of truths as actual to the apprehension of them as necessary. (6) This successive and various progress from fact to theory constitutes the history of science; (7) and this progress, though always leading us nearer to that central unity of which both the idea and the fact are emanations, can never lead us to that point, nor to any measurable proximity to it, or definite comprehension of its place and nature.³⁴

But all the hesitation appears to have gone away from Whewell's mind when he wrote the theological sections of *Philosophy of Discovery* and *Plurality of Worlds*. In these sections the claim is made that the central unity is indeed reached, and reached in precisely those intuitions which are apprehensions of necessary truth.

Kant's transcendental deduction would not serve for Whewell as a solution of his particular form of the problem of establishing epistemic correlations between conceptual forms and particular empirical perceptions. Nor would the argument of the two memoirs on the fundamental antithesis to the effect that ideas and things are in fact inseparable, however much they may be separated for purposes of philosophical understanding, completely solve his problem of justifying the intuited forms as forms that actually form some empirically real objects that exist independently of the forms. For Whewell was perceptive enough as a philosopher to realize that if the terms of the fundamental antithesis (for example, "ideas" and "things") are to be derivable from one another, or reducible to one another (as the argument of the two memoirs

³³ PW, pp. 276-278. This quoted material is from Whewell's printer's copy of PW, dated 1853 and bearing the catalogue number ADV.C.16.27 in the Wren Library, Trinity College, Cambridge. This copy of PW contains five chapters that Whewell had printed but then later deleted from the published edition. Apparently he suppressed this material at the suggestion of Sir James Stephen who complained that these five chapters on metaphysics could not appeal to the same audience to which the main body of the book was addressed. In a letter to Sidgwick (MSS. in Wren Library, Trinity College, dated June 8, 1854) Whewell explains why he suppressed these chapters: "In the first printing of my essay I had pursued my speculations about the Divine Mind a good deal further than in the published book. I suppressed what I had printed because I thought that the greater part of my readers would be repelled by what they would call metaphysics; but if you could find time to read my cancelled pages, I think they would interest you; and I should be very much pleased to hear your opinion about my speculations." It should be pointed out that Whewell did not suppress all theological material from the essay on *Plurality of Worlds*. For example, in the finally printed version of PW, Whewell argues on pp. 109-110 that the Ideas are both epistemological conditions of human knowledge and ontological conditions of the creation of the world by God. The five suppressed chapters express this position much more fully, but the point is that it was Whewell's published position. It seems clear, then, that Whewell was yielding to editorial advice in the act of suppressing these five chapters, and that he had not come to disagree with what he wrote there. That these five chapters do represent a metaphysical position that Whewell accepted toward the end of his life is amply confirmed by the fact that Chapters XXX-XXXI in PD essentially reproduce this once suppressed metaphysical material. In addition there is also some discussion of the relation between the Divine and the human mind closely resembling a discussion of this topic in one of the suppressed chapters of PW in an uncatalogued manuscript notebook entitled *Philosophy and Theology* dated August 27, 1851 (Wren Library).

³⁴ FAII, "Additional Note to two Memoirs 'On the Fundamental Antithesis of Philosophy'."

contents but does not prove), there must be some principle of unity that subsumes both terms in the pairs that Whewell lists. There must be some "third man" (the problem is a variation on Kant's schematism problem) which is like both Ideas and Things, or "... some central unity of which both the idea and the fact are emanations. . . ." Only if this third man can be located can we in point of ontological fact be convinced that what we intuit as necessary is a form of *something*, rather than a mere logical phantom, a *Kantian* metaphysical idea. And what Whewell discovered was that he could not locate this third man in the region of a Kantian epistemology, nor in the region of a mixed epistemology-metaphysics as is exhibited in the two memoirs, and shows definite marks of the influence of the German romanticists, Fichte, Schelling, and Hegel.

So Whewell looked for the third man in the region of that fairly orthodox and straightforward Christian theism that had rested quietly behind the scenes awaiting eventual re-introduction.³⁵ For the third man that Whewell finally discovered (though perhaps "discovered" must here be regarded as a euphemism) was God, from whom emanated both ideas and things, and in whom both ideas and things are identical. The non-Kantian, seventeenth-century rationalist ring of Whewell's final theological deduction of the Ideas is unmistakable:

Why does the constitution of the world, as an object, correspond with the constitution of the mind, as a mind? First; because they are works of the same Maker; but that is not the whole reason: further, because the constitution of the world is marked with the Thoughts of the Divine Mind, and the human mind is, in part, a sharer in the Thoughts of the Divine Mind. And thus, we not only see that objects do exist in space and under *spatial* relations; but we feel that they must so exist, and that spatial relations are necessary truths. (PW, p. 282)

In the final analysis, then, science is the right interpretation of nature, because the forms of scientific knowledge are also real conditions of the nature of things. Given this final outcome of his system, it seems apparent that Whewell wanted something that would steer a middle course between the formalism of Kant's critical idealism and the empty conventionalism of which he had

been accused by some of his critics. This means that we may stress, as Whewell himself certainly did, the imaginative and constructive aspects of scientific work, and we may even point out, what Whewell also certainly did, the epistemological presuppositions of finished systems of science. But as science for Whewell was not a self-justifying and self-critical epistemological enterprise, so neither the methodology of the discovery of scientific hypotheses nor presupposition-hunting are self-justifying and self-critical. What supplies the justification is certainly what Whewell called "metaphysics." Whewell's final metaphysics is rather clearly a realistic one and it pretends to rule out both critical idealism and conventionalism.

In a paper devoted to exposition of Whewell's views (a difficult enough project given Whewell's own peculiarities in setting forth his doctrines) criticism of his system is perhaps out of place. But I think it must be mentioned that Whewell's final solution of the "ultimate problem" is unsatisfactory, because his final theology of science is unsatisfactory in at least two ways. He offers no substantial justification of the epistemological status of theological claims to truth (which is after all only another species of the same problem that worried him relative to intuition of necessary truths), and he does not work out the details of his theology in anything like a complete philosophical way. Without supplying adequate ways of making up these two defects, it is not clear why a philosophy of science must look to theology for its completion, nor is it clear how theology can mediate between conceptual systems in the sciences and perceived facts. But for all the defectiveness of his final system, Whewell did see the limits of Kant's theory of knowledge, and he saw more clearly than any other nineteenth-century philosopher of science had seen the epistemological need for abstract conceptual systems and empirical facts, and the peculiar philosophical problems that result from attempting to harmonize these two heterogeneous elements.

VI. CONCLUDING REMARKS

There is, I think, a moral to this story. The influences on Whewell's theory of science were many and strong. On the one hand the critical philosophy of Kant made its mark. On the other

³⁵ Of course in one sense the theology had not rested at all, for one clear statement of it in the *Bridgewater Treatise*, which appeared unchanged in seven editions between 1833 and 1864, was always available. But Whewell's theology in PW and PD differs from that in BT. The differences, however, must await discussion in another place.

hand, surely never to be forgotten in the case of Whewell, the history of the inductive sciences, as he understood that history, made its mark too. Thus he could accept Kant's general position on categorial knowing, but only in a context that made the discovery of more categories possible, and made possible also the introduction of that kind of metaphysics that Kant's criticism attempted to rule out.³⁶ Whewell's system shows the scars and open wounds of attempting to develop a body of philosophical doctrine on science out of materials partly logical and epistemological (Kant), and partly historical, though Whewell's history itself operates in the context of more or less explicit philosophical presuppositions about the nature of science. The entire problem of justifying a doctrine of the progressive intuition of necessary truth might never have arisen if he could have kept separate in his thought the quite different processes of giving a historical account and attempting a logical or epistemological analysis, processes which he was more likely to confuse than recognize as different. Thus he could not appreciate fully Mansel's telling criticism of his analysis of necessity, a criticism based upon a set of careful logical distinctions to which Whewell's more historically oriented mind was largely indifferent.³⁷ But it must be said on his behalf that Whewell did make the first grand effort to harmonize the history and philosophy of science, and that this effort took the form of trying to work out a coherent philosophy of *discovery*, rather than a philosophy of *proof*. It was a mark of his times to confuse logic with *something*—most of his contemporaries confused logic with psychology—and his confusion of it with history is one more form of that psychologism in logic and theory of knowledge that had to await the demolishing criticism of Frege, Husserl, and Bradley to be finally put aside.

In the end, however, both Kant and the history of the inductive sciences together had to submit to the third great influence on Whewell's thought: an unquestioned, literally ontological, Christian theism. And after all, when the really crucial questions arose in his system, what else was available to

him? Mill's inductivism was no option—Whewell had already criticized its crucial defects. Nor was a conventionalistic philosophy of science a live option—Whewell saw the philosophical problems too clearly to wash them all away as he thought that Comte had done. The new mathematical logic that his friends and colleagues (DeMorgan, Boole) were beginning to develop might have provided eventually a way of harmonizing the philosophy and the history of science (as some contemporary philosophers have done, rather paradoxically by drawing a sharp line of demarcation between them and then concerning themselves only with analysis of the logical structure of sciences), but the slim avenue of escape provided by this new tool could not have interested one accustomed to travel on grand, wide philosophical avenues. So there was only the Christian theism that Whewell had proclaimed in Trinity College Chapel and defended against the infidels in several printed works to fall back on. And the fall was not really a fall from either historical or philosophical grace; for many of the details of his *History* and *Philosophy* can still be reckoned among the enlightened and virtuous acts of nineteenth-century scholarly literature.

REFERENCES TO WHEWELL'S WORKS CITED

- BT *Astronomy and General Physics Considered with Reference to Natural Theology* (London, 1833).
- DMH "Demonstration that All Matter is Heavy," *Transactions of the Cambridge Philosophical Society*, vol. 7, pt. II (1841). Reprinted as an appendix in PD and PIS.
- EM *The Elements of Morality, including Polity* (London, 1845), 2 vols.
- FA "On the Fundamental Antithesis of Philosophy," *Transactions of the Cambridge Philosophical Society*, vol. 7, pt. II (1844). Reprinted as an appendix in PD and PIS.
- FAII "Second Memoir on the Fundamental Antithesis of Philosophy," *Transactions of the Cambridge Philosophical Society*, vol. 7, pt. V (1848).

³⁶ One of Whewell's reviewers praises the Kantian spirit of his *Philosophy* and the merits of its anti-empiricist effects. But he warns Whewell not to take Kant too seriously, since his "ultra-rationalism" in religion would corrupt Cambridge. "Whewell's 'Philosophy of the Inductive Sciences,'" *Dublin University Magazine*, no. 98 (Feb. 1841). I think that Whewell was more prone to accept this sort of remark than his earlier commentators seem to have realized. The religious and moralistic background of Whewell's philosophy has yet to be studied.

³⁷ See H. L. Mansel, *Prolegomena Logica, An Inquiry into the Psychological Character of Logical Processes* (Oxford, 1851), pp. 189–293; and "The Limits of Demonstrative Science Considered in a Letter to the Rev. William Whewell, D.D." (Oxford, 1853). Also see Whewell, "A Letter to the Author of *Prolegomena Logica* by the author of the *History and Philosophy of the Inductive Sciences*" (Sept. 20, 1852), and PD, pp. 335–339.

- | | | | |
|-----|--|-----|---|
| HIS | <i>History of the Inductive Sciences</i> , 1st ed. (London, 1837), 3 vols. | PEU | <i>On the Principles of English University Education</i> (London, 1837). |
| HSI | <i>The History of Scientific Ideas</i> (London, 1858), 2 vols. Part I of the 3d. ed. of PIS. | PIS | <i>The Philosophy of the Inductive Sciences</i> , 2d. ed. (London, 1847), 2 vols. |
| IM | <i>Of Induction, with Especial Reference to Mr. Mill's System of Logic</i> (London, 1849). Reprinted as part of PD. | PTI | "Of the Platonic Theory of Ideas," <i>Transactions of the Cambridge Philosophical Society</i> , vol. 10, pt. I (1857). |
| ME | <i>Mechanical Euclid</i> (Cambridge, 1837). One section, "Remarks on Mathematical Reasoning and on the Logic of Induction," reprinted as an appendix in PIS, 2d. ed. | PW | <i>Of the Plurality of Worlds</i> (London, 1853). Also copy (dated also 1853) containing 5 chapters printed but cancelled in the published work (Wren Library, Trinity College, Cambridge). |
| NOR | <i>Novum Organon Renovatum</i> (London, 1858). Part II of the 3d. ed. of PIS. | TD | <i>On the Motion of Points Constrained and Resisted, and on the Motion of a Rigid Body. The Second Part of a new edition of A Treatise on Dynamics</i> (Cambridge, 1834). |
| NTM | "On the Nature of the Truth of the Laws of Motion," <i>Transactions of the Cambridge Philosophical Society</i> , vol. 5, pt. II (1834). Reprinted as an appendix in PIS. | TM | <i>An Elementary Treatise on Mechanics</i> , 1st ed. (Cambridge, 1819), and 5th ed. (Cambridge, 1836). |
| PD | <i>On the Philosophy of Discovery</i> (London, 1860), pt. III of the 3d. ed. of PIS. | TSM | <i>Thoughts on the Study of Mathematics as Part of a Liberal Education</i> (Cambridge, 1835). |

Middlesex College,
University of Western Ontario

II. RECENT WORK IN AESTHETICS

JOSEPH MARGOLIS

RECENT contributions to the Anglo-American literature in aesthetics appear chiefly, though not exclusively, in the journals. And aesthetics itself concerns a very wide range of rather loosely related issues. It is convenient, therefore, in canvassing the field, to bring together the most prominent or most instructive contributions on particular topics rather than to attempt to summarize systematically the work of particular authors.

An earlier vogue in aesthetics, of about the time of the appearance of Elton's anthology [22],* was distinctly concerned to oppose idealism and to exploit in a very general manner the new analytic currents informed principally by the inquiries of Wittgenstein. Since that time, idealism has somewhat declined as the inevitable opponent, with the withering away of the expression "theory of art," and analytic contributions have become more detailed in inquiries into critical language but have remained as eclectic as ever in their preference of philosophical models.

Of all the questions that fall within the range of aesthetics, those that concern the nature of criticism lend themselves to the most systematic ordering. Here, the principal topics include the description, interpretation, appreciation, and evaluation of works of art, the analysis of the characteristic predicates that are employed in critical remarks either by professionals or amateurs, and the very variety of functions that critical exchange is designed to serve—including for instance instruction and the exhibition of taste, the analysis and comparison of works of art. We may also include here questions regarding the relevance or lack of relevance of comments about perceivable properties of a work of art, or about the conditions of production of a given work, or biographical information about the artist, or about the sorts of responses of an aesthetic percipient. There is no way to organize the relevant discussion simply, but the issues themselves are clearly interrelated.

From the time of the appearance of Elton's collection, one may fairly say that the analysis of

the language of criticism has been pursued against the background of Margaret Macdonald's [54] and Arnold Isenberg's [43] accounts. Both of these authors, each in his own way, have been intent on insisting that supporting evidence for a critical "verdict" about some work of art could not, in principle, be supplied so as to oblige another to accept that verdict. Neither wished to deny, in some sense, the relevance of supporting reasons for a judgment rendered, but Isenberg was inclined to argue that a critical judgment is at bottom a matter of taste or feeling and Macdonald argued that reasons supplied are rather more like the detailing or "presenting" of what it is one appreciates in a given work than the provision of decisive evidence. Both rather misleadingly speak of "verdicts," though it is clear that their accounts do not intend the term in its usual sense. The common tendency of their discussion is relatively widespread. For example, it has been reinforced, particularly where emotional or expressive qualities of works of art are in question, as by John Hospers [35] and Henry Aiken [1], so that it appears to infect the description of works of art. And, to the extent that the interpretation of works of art is construed in terms of preferences of taste, as in C. L. Stevenson's recent discussion [81], critical judgments not overtly valuational in nature have been construed as not confirmable in the usual way in which matters of fact are (following, in Stevenson's case, the well-known account of value judgments as inherently persuasive) [82], [83].

The tendency of discussions of these sorts (and they are by no means in agreement with one another) falls very nicely in accord with that first wave of reaction to G. E. Moore's account of good as a non-natural property, in the sense that valuationally significant predicates in the aesthetic setting are not admitted to designate the properties of any work of art. The Naturalistic Fallacy has gradually lost its fearsome aspect and theorists in aesthetics (as well as in ethics and allied fields) have found it possible to reconsider the ascription of these predicates as objectively confirmable. The

* For references of this sort see the "List of Works Discussed" given at the end of the paper.

traditional view of public canons for evaluation is upheld for example in Monroe Beardsley's *Aesthetics* [6], where it is maintained that criteria corresponding to such broad-gauge judgments as of "unity," "harmony," and "balance" may be explicitly formulated and employed; and even J. A. Passmore's critical paper [6] shows that that author is prepared to support, contrary to the reader's first impression, objective valuational criteria, though not perhaps of the same gauge that Beardsley is confident may be provided. The analysis of relevant valuational usage in P. H. Nowell-Smith's *Ethics* [66] marks more definitely the beginning of the recovery of objectively attributable predicates both in ethics and aesthetics, with a fuller appreciation of the variety of conditions on which these may be appropriately used. And the most sustained and distinctive account of narrowly aesthetic concepts has been supplied by Frank Sibley [78]. Sibley maintains that characteristically aesthetic concepts, like "garish," "dainty," and the like, cannot be segregated into descriptive and valuational components and are in fact not condition-governed at all, as are, in a variety of ways which he details, other characterizing predicates. His account is original and challenging and has provoked an exchange [76], [77] and comments. The puzzle remains that aesthetic qualities are taken to be perceivable, to be complex in the sense in which they depend on non-aesthetic qualities, but simple in the sense in which criteria for their occurrence cannot be specified. In this respect, Isabel Hungerland [39] has relevantly argued that Sibley's principal specimen terms do not allow, in application, for a contrast between what only seems to be so and what is really so. The upshot is that the question remains whether (and in what sense) one may be said to perceive that an object may be characterized in the relevant ways as opposed to one's being entitled to attribute to an object a relevant predicate on the basis of one's own taste. Also, the question may be raised as to whether the so-called aesthetic concepts (in a reasonably generous sense of "aesthetic") do show a common (and this particular) logical pattern; Sibley himself concedes that there may be exceptions, and the analysis of important categories like the baroque and the tragic do not seem to be readily analyzable along the lines suggested. The baroque, in fact, has received considerable attention in recent years [26], [32].

The discussion of value judgments has, since Moore, increasingly turned away from an almost

exclusive concern with the use of "good" to predicates that have significant descriptive force. That valuational terms are used in other than descriptive ways, for instance for commending purposes, is part of the permanent contribution of analysts like Nowell-Smith [66] and R. M. Hare [28] who have responded to Moore's challenge. But, more pointedly for aesthetics, the early confidence of a commentator like Helen Knight [46], who holds that there are formulable sets of criteria for all the characteristic uses of "good" in such phrases as "good show-dog," "good novel," "good tennis racquet," "good landscape," does not seem to be readily supported. The objectivity of such judgments need not be in question in challenging the view that there are necessary and sufficient conditions on which these judgments rest that can be antecedently provided. Criticism characteristically does not, and as William Kennick [45] argues, summarizing the accumulating evidence, need not work by applying rules to cases. A variety of analysts, chiefly in the ethical and legal spheres (for instance, Kurt Baier [5], Philippa Foot [23], H. L. A. Hart [29]), have tried to recover the objectivity of particular valuational predicates.

It is clear, of course, that aesthetic judgments characteristically, though not exclusively, depend on one's taste in an intimate way. The Kantian division of aesthetic and moral values, recently revived in an ingenious form by Stuart Hampshire [27], may be shown to be untenable. Aesthetic and moral judgments may exhibit the same logical properties and, within each domain, there may well appear logically distinct types of judgment. The most general account showing that aesthetic and moral judgments cannot be viewed as merely separate genera of judgment may be found in J. O. Urmson's account [88]. And Joseph Margolis [59] has sought to show that value judgments are of at least two distinct logical types (ranging across both domains)—"findings," which logically behave the way factual judgments do, except that the categories employed are valuationally significant; and "appreciative judgments," which depend on our particular tastes (both are supportable and, though appreciative judgments are weaker than findings, they are not, for that reason, less objective). Generally speaking, the trend in value theory is very decidedly in favor of exploring the actual grounds, the relatively informal grounds, on which relevant judgments are rendered. In a word, what has come to be called

cognitivism in value theory has been revived compatibly with the positive contribution of Moore's early statements.

Special attention has been paid to so-called expressive qualities in a work of art. Traditionally, with the Idealists, these have been linked to an account of artistic creativity; alternatively, following the example of Santayana, they have been construed as a projection of the response of the aesthetic percipient. Varieties of these views may be found for example in Stephen Pepper [68], and L. A. Reid [73]. At the present time, along lines substantially in accord with Vincent Tomas' account [85], expressive qualities are construed as confirmable in a public way and substantially freed from an account of creativity, even if (following Nowell-Smith) dispositions to affective response are relevant. The view has been somewhat opposed by Hospers [35] and Aiken [1], both of whom are inclined to view the relevant comments as personal and subjective; also by Beardsley, but (as in "The Affective Fallacy" [97]) for the altogether different reason that aesthetic considerations are thought to preclude attention to affective responses. What may be said in general, conceding a variety of uses of expressive predicates, is that there is a substantial range of such predicates used in a genuinely characterizing way that do not behave in a manner significantly different from that of otherwise quite typical aesthetic concepts.

Another cluster of issues centers on the interpretive efforts of critics. Stevenson [81], as already noted, conflates interpretive efforts with the expression of personal taste and the influence of others. He does not concern himself at all with the question of professional criteria relatively free of personal preference or with the distinction between evaluating a work of art and evaluating an interpretation of a work of art. The argument is, to this extent, unconvincing. A number of commentators have explored the issue of interpretation in ways that seem reasonably free of Stevenson's reductive view; and Stevenson himself, notably in *Ethics and Language* [82], had been obliged to face a corresponding difficulty in construing technical criteria of truth and validity in science and logic as somehow not "normative." The weakness of the present argument corresponds with that symptomatic difficulty. Granting relatively clear-cut professional canons (admittedly somewhat informal), the discussion has moved along two distinct lines. Some theorists, notably Beardsley [6] and Pepper [69], have held that defensible interpretations of a work

of art must all be ideally convergent, even if (as Beardsley would insist) that ideal interpretation cannot be supplied. Margolis has argued [60], to the contrary, that, given a work of art whose describable properties are noted, it is possible in principle, though not always fruitful in practice, to provide defensibly plural, non-converging, even incompatible interpretations of a given work of art. Suggestions along this line may be found also in Macdonald [54], [57].

Finally, attention has been drawn, notably by Paul Ziff [98], to the varieties of remarks, all bearing on the discriminable properties of a work of art, that may or may not in different ways be relevant to instructing another in what one may appreciate in a given work. Ziff introduces the useful concept of "aspection" and stresses the point that one cannot offer reasons to another for valuing a work that are avowedly personal, though such reasons will be personal. The thesis accords generally with the view that one's personal tastes are never, as such, reasons that weigh with another's judgment and appreciation. The confusion of taste and judgment (even appreciative judgment) may be found in Bernard C. Heyl [33], [34] and colors to a degree Isenberg's account [43]. It may be useful to mention here that detailed analyses of appreciative exchange are comparatively rare.

There is, to consider a second run of issues, a cluster of classic questions about the literary arts, which, though they do not have any clear unity, are readily collected. The principal topics concern fiction and metaphor. Two symposia in the *Proceedings of the Aristotelian Society* [10], [55] have been particularly instructive about fiction. There we may see that an earlier concern had been the elimination of fictional entities from the real world, what may fairly be regarded as an extension of the attack on idealism launched by Moore and Bertrand Russell. The pivotal issue had been reference to characters in fiction; under the influence of linguistic analysis, discussion gradually settled on the fictional use of sentences. Discussants, therefore, all subscribe to the view, emphatically put by R. B. Braithwaite [10], that the required distinctions must be detailed within "one world," but there is considerable divergence about the logical properties of fictional sentences. Gilbert Ryle [10] had been strenuously criticized by Moore [10] for thinking that fictional sentences might, by a fluke of chance, turn out to be true. But Ryle himself had held, correctly though perhaps not

consistently, that fictional characterizations employ only "pseudo-predicates," and Moore's correction was itself decidedly ambiguous and even, in a way, a repetition of Ryle's mistake. For Moore uses locutions that cannot escape being construed as referring to fictional entities and he holds that what is given in the story is "false." Furthermore, Moore finds it necessary to cast the concept of a story as a set of statements that the author makes, and this leads him to speculate about authors' biographies in an altogether unnecessary way. The best account in these symposia is undoubtedly that of Margaret Macdonald [55] who carefully distinguishes fiction from lies and falsehoods, construing it as an implicit conspiracy between reader and author in terms of which questions of truth and falsity are waived. A position very much like that of Ryle's, but stated even more positively, is advanced by Beardsley [6] who believes that a story may actually turn out to be true. But his account fails to distinguish, for instance, between the logical function of sentences used to tell a story and the reasons for which a court of law may choose to treat an account as a libel rather than as a fiction. The puzzles of reference in fiction may be resolved by distinguishing that use of sentences (which Macdonald discusses) by which we are to imagine a certain world to exist, from those uses, given the imaginary world, that correspond with all the uses eligible in the real world: the problem of reference is thereby reconciled with current discussions of the referring use of language.

But the problem of truth in fiction, and in literature in general, and even in art in general, has remained interesting. The most promising adjustment, in the setting of fiction, has involved an application of Moore's well-known account of a certain sort of implication. The concept has been explored, in a non-aesthetic context, most systematically by Isabel Hungerland [42]. And, in the aesthetic context, it has been applied to fiction notably by Morris Weitz [95] and John Hospers [36] (who, in this respect, departs from his earlier views [37]). The difficulties are of two sorts. For one, it is necessary to distinguish verisimilitude in a novel or fiction from the (so-called "contextually implied") truth of a fiction. Weitz's discussion shows the ease with which one may slide from a comparison of a fictional character or scene with elements of the real world (verisimilitude) to a detailing of propositions allegedly "implied" by the fiction itself. And for another, one must distinguish the view that a work of art is a clue to the artist's

convictions from that which holds that the work of art somehow presents or "implies" statements that otherwise might have been asserted. Hospers' discussion shows the difficulty of distinguishing between our making inferences about the author's views from his fiction to discriminating those propositions that are "contextually implied" by the fiction itself. The real problem lying behind these discussions remains with the informality of the type of implication identified.

The principal questions regarding poetry concern metaphor and symbol, but symbol lends itself to a larger discussion involving all the arts. The two are confused as differences of degree rather than of kind by such literary theorists as Cleanth Brooks [11]. Reasonable grounds for distinguishing the two are discussed in perhaps the most sustained account by Isabel Hungerland [40]. The single most influential paper on metaphor is undoubtedly that of Max Black [8]. Paraphrasability is generally regarded as the central issue. Beardsley [6], for instance, regards metaphor as a "significant attribution" which, apart from its ornamental features, may be paraphrased by singling out the appropriate connotation and supplying attributions that may be truly or falsely ascribed to the relevant object. Beardsley is inclined, therefore, to associate metaphor with assertion and statement. Metaphor thereby becomes more or less of a puzzle to be solved with ingenuity and, in principle, paraphrasable. Paul Henle [30] develops an iconic theory of metaphor, which concerns both paraphrasability and the question of the basis for a metaphoric invention. He is unable to distinguish metaphor satisfactorily from catachresis (which at least Gustaf Stern, in his well-known account [80], had warned about). Also, though he employs C. S. Peirce's category of the iconic, Henle considerably strains Peirce's view of the limits of iconic similarity. I. A. Richards [74] had already considered grounds other than similarity as a basis for metaphor, and the ubiquity of resemblances of all sorts suggests (the argument goes against Andrew Ushenko [89] as well) that mere resemblance can hardly explain metaphor. Richards, incidentally, departing from Aristotle, had held metaphor to be an essential feature of language. He has been followed in this by a number of writers [7], [24], but most discussants treat metaphor correctly as logically dependent on non-figurative uses of language. In fact, the entire enterprise of paraphrase requires such a view. Black adopts a moderate position, allowing for paraphrasability for certain sorts of

metaphors but holding, in his most characteristic view—in what he terms “the interaction view of metaphor”—that paraphrase would involve a loss of “cognitive” content, that metaphor provides a new insight into things. He views metaphor as a sort of “filter” through which the world may be seen in a fresh light but which antecedent language has not yet reached; “successful” metaphors will then decay in the manner of catachresis. Margolis [59] has argued that paraphrase is not so much inadequate as irrelevant to metaphor, holding that metaphor is primarily a “game” of language by which we deliberately deform things and the use of characterizing terms so that we may play with things as if they were the same or similar knowing that they are not; the purposes of using metaphor and of paraphrasing language are at odds with one another.

About poetry in general, it may at least be said that I. A. Richards’ attempt [76] to assign a distinctive logical or linguistic function to poetry has been largely abandoned. Some of the most inventive speculations about alternative types of poetry (as well as of other types of literature and of art) have been supplied by Susanne Langer [48]. But it seems, for instance, inappropriate to speak of poetry as contrasted with fiction, if for no other reason than fictional tales are often told poetically. This suggests considerations discouraging any simple generalization upon the variety of things falling under the head of poetry. The key distinction regarding symbols, as for instance discussed by Isabel Hungerland [40], concerns the need, in the setting of critical appreciation, to construe actions, objects, remarks, and the like as symbolizing something not otherwise presented, whose recognition by the critic (or reader, viewer, listener—for symbols, unlike metaphors, are not peculiarly literary devices) permits the entire structure of the work to be grasped and formulated. In short, the appeal to symbols concerns the clarification of works of art whose explicit structure is otherwise puzzling and difficult to articulate. The question of the appropriateness of construing materials symbolically returns us once again to the procedures of critical interpretation and dramatizes the problem of identifying and neutrally describing works open to alternative interpretations [59], [63].

There have appeared, also, in recent aesthetics, two quite strenuously debated but rather special issues. One concerns the so-called Intentional Fallacy; the other, the question of the definition of

art. The first was prompted by the appearance of William Wimsatt and Monroe Beardsley’s well-known paper [97]; the second, by the appearance of Morris Weitz’ discussion of the problem of definition [96]. Regarding the first, Wimsatt and Beardsley had held that the artist’s intentions were neither available nor aesthetically relevant (nor desirable) in the critical appreciation of literature (or, one is inclined to suppose, in the other arts). Their discussion is somewhat spoiled by altogether too mentalistic a view of intentions. The paper, for instance, has been interestingly criticized by Theodore Redpath [72] who suggests reasonably that an artist’s intention may be discovered without going outside the work itself. Also, the aesthetic relevance of an artist’s intentions has been argued for by Hungerland [41], Aiken [2], and Eliseo Vivas [90], among others. Furthermore, a careful reading of the original article will show that the artist’s intentions are actually not taken to be necessarily unavailable or even irrelevant; what the authors insist on is that the artist’s intention cannot be a privileged criterion by which to decide what a work means. The controversy is somewhat additionally complicated by the fact that Beardsley [6] has since construed the problem as that of our being actually able to fix the artist’s intention and of determining the extent to which a work produced falls short of such intention. But this appears to undermine very nearly the original thesis.

Weitz’s discussion depends on his attempted application of Wittgenstein’s well-known account of “family resemblances” to the concept of art. He argues that it is, in principle, not possible to define any “empirical” concept in terms of necessary and sufficient conditions. But there are difficulties with his view. For one thing, Wittgenstein in the passage in question does not preclude definition, but only shows that concepts resting on “family resemblances” are readily usable. For another, Weitz finds that Aristotle’s definition of tragedy is false, though not inappropriately formulated. And finally and most important, Weitz does not satisfactorily distinguish between the logical job of defining a concept and that of extending the use of a concept: he appears to foreclose on the possibility of the first by appealing to the implications of the second. He has, in these respects, been criticized by Margolis [61]. But the appearance of his paper has undoubtedly drawn attention to ulterior uses of definition for the presentation of otherwise eligible aesthetic theories.

There are also several even more specialized

controversies that are of some interest. For one, there is the matter of Susanne Langer's theory of symbolic forms. First formulated in *Philosophy in a New Key* [49], it was decisively criticized by Ernest Nagel [64] in a review. Langer had expanded her theory without satisfactorily meeting Nagel's objection, and in her most recent collections of essays [50], [51], she has more or less abandoned the designation "symbolic" for the distinction she has in mind. There has also been a lesser exchange regarding Stephen Pepper's concept of a work of art as a "nest of objects" [69], [70]. Pepper's principal critic in this regard has been Donald Henze [31]. Criticism centers on the oddity of construing a work of art produced by an artist as a set of plural objects; the issue has implications for the objectivity and relevance of critical and appreciative remarks.

Some of the most energetic controversies, in fact, have centered around some proposed definition of a work of art. Mention may be made for example of Vivas' [90], [91] challenging the neo-Aristotelian definitions of the Chicago School [15] and of Macdonald's criticism [56] of Weitz's [95] organismic theory of art (which he has acknowledged). Finally, the expression "theory of art" has been exploded, at least in the classic form in which Bernard Bosanquet [9] had presented it—in the form in which one supposes that the artist's feelings have somehow gotten into the work of art itself. The continuing, but somewhat dwindling, criticism of this issue shows fairly clearly that it is something of a vestige. The best-known discussions of its weaknesses appear in Tomas [85] and Hospers [35], [38]. The expression theory, of course, is simply the philosophically most interesting version of the various theories of artistic creativity. These latter are undoubtedly suggestive but are difficult to manage in more than an impressionistic manner, which may be readily seen by reviewing such authors as Benedetto Croce [16], John Dewey [18], C. J. Ducasse [21], and R. G. Collingwood [14], the latter of whom in many ways is the most explicit. Freudian and Marxist theories are very nearly useless—though they are obviously not irrelevant for criticism [5], [13], [47]—at least for the reason (almost explicitly conceded by Freud himself) that they are not focused on the artist's craft. Possibly the most sustained and suggestive (but surely curious and unmanageable) account is given by Jacques Maritain [58] who appears to combine Bergson and St. Thomas. There are also accounts of special aspects of the question of

creativity in Milton Nahm [65], Paul Weiss [94], and Tomas, who has collected a small anthology of relevant papers [86].

Interest has also centered on the nature of aesthetic experience and associated concepts. Earlier accounts had tended to draw attention to distinctive discriminations made possible by aesthetic orientation itself. L. A. Reid [73], for instance, had argued that the expressive qualities of a work of art are simply projections of the response of an aesthetic percipient. Similar, though not entirely unequivocal, views have been put forward by Pepper [68] who is inclined to hold that there are properties of a work of art not discriminable unless one takes an aesthetic attitude to the work in question; Ducasse [21] also holds that there are distinctive qualities that can only be "ecpathized," that is, discriminated if one adopts the aesthetic attitude. All such doctrines are closely akin to (though substantively distinct from) the views of such theorists as Edward Bullough [12] and the empathists [52], [53]. Bullough's thesis has been recently revived [17] but the original difficulties of his doctrine have not been resolved—in particular, that the negative and positive scales of "distancing" have nothing in common, which upsets the "antinomy of distance," and that the doctrine is merely designed to favor certain special aesthetic values (as with the empathists as well). Tomas [87] has attempted to hold that aesthetic discrimination attends only to "appearances," in an effort to assign a distinctive object to the aesthetic concern. His account has been criticized by Sibley [79], for instance, who points to the anomalies to which this leads. Beardsley [6] had committed himself to a view similar to Tomas'. Possibly the most ambitious effort to distinguish aesthetic perception from that sort of perception that obtains in science is offered by Virgil Aldrich [3] who contrasts "basic imagination" with "observation," which he takes to be fundamentally different aspects of perception. The problem remains that, admitting strikingly subtle perceptual discriminations in the aesthetic setting, it is difficult to see why these discriminations theoretically require a distinct "mode" of perception; Aldrich's illustrations [4] do not seem to force us to hold the view he advances. Margolis [62] has argued that there is no distinctive perceptual "mode" corresponding to aesthetic interest, though perception, imagination, memory, association, and the like play a number of distinct roles in the appreciation of a work of art. A similar view is

advanced by George Dickie [19]. Psychological explorations of aesthetic experience, it has been argued both by Dickie [20] and Morgan [63], have not been particularly productive. But the concept of the "perception" of a work of art is misleadingly unified and a whole host of special problems arise regarding all the conceivable "ingredients" of such perception. For instance, special questions arise even for visual perception, as in representational art—which have been discussed by Isenberg [44], Ziff [99], and Stevenson [84]. The question of the appropriate sense in which one may be said to "perceive" literature had already been broached by David Prall [71], and has been more recently considered by Sidney Zink [100]. And of course, as we have already seen, the discrimination of emotional and expressive qualities in a work of art

has raised its own distinctive questions. The search for distinctive values, perceptions, and experiences in the aesthetic setting—instructively, for instance, in such different writers as Dewey [18] and Vivas [90], [92]—may, not unreasonably, be regarded as vestigial remains of the original Kantian division between the aesthetic and the moral. One may say that, bit by bit, the basis for a clear and simple logical contrast between these two domains has been whittled away.

These, then, are the principal issues that have occupied the attention of aestheticians in recent years. They do not exhibit a ready unity, and very likely, one cannot avoid in a survey of this sort favoring one set of issues rather than another. A selected bibliography is appended to offset, if possible, the inevitable incompleteness of the account.

LIST OF WORKS DISCUSSED

1. AIKEN, Henry "Art as Expression and Surface," *Journal of Aesthetics and Art Criticism*, vol. 4 (1945).
2. ——"The Aesthetic Relevance of Artists' Intentions," *Journal of Philosophy*, vol. 52 (1955).
3. ALDRICH, Virgil "Picture Space," *Philosophical Review*, vol. 67 (1958).
4. ——"Philosophy of Art (Englewood Cliffs, N.J., 1963).
5. BAIER, Kurt *The Moral Point of View* (Ithaca, N.Y., 1958).
6. BEARDSLEY, Monroe *Aesthetics* (New York, 1958).
7. BERGGREN, Douglas "The Use and Abuse of Metaphor," *Review of Metaphysics*, vol. 16 (1962-63).
8. BLACK, Max "Metaphor," *Proceedings of the Aristotelian Society*, vol. 55 (1954-55).
9. BOSANQUET, Bernard *Three Lectures on Aesthetic* (London, 1915).
10. BRAITHWAITE, R. B., RYLE, Gilbert, and MOORE, G. E. "Imaginary Objects" (symposium), *Proceedings of the Aristotelian Society, Supplementary Volume XII* (1933).
11. BROOKS, Cleanth and WARREN, Robert Penn *Modern Rhetoric* (New York, 1949).
12. BULLOUGH, Edward "Psychical Distance as a Factor in Art and an Aesthetic Principle," *British Journal of Psychology*, vol. 5 (1912-13).
13. CAUDWELL, Christopher *Illusion and Reality* (New York, 1937).
14. COLLINGWOOD, C. G. *The Principles of Art* (London, 1938).
15. CRANE, Ronald S. (ed.) *Critics and Criticism, Ancient and Modern* (Chicago, 1954).
16. GROCE, Benedetto *Aesthetic* (New York, 1909).
17. DAWSON, Sheila "Distancing as an Aesthetic Principle," *Australasian Journal of Philosophy*, vol. 39 (1961).
18. DEWEY, John *Art as Experience* (New York, 1934).
19. DICKIE, George "The Myth of the Aesthetic Attitude," *American Philosophical Quarterly*, vol. 1 (1964).
20. ——"Is Psychology Relevant to Aesthetics?" *Philosophical Review*, vol. 71 (1962).
21. DUCASSE, C. J. *The Philosophy of Art* (New York, 1929).
22. ELTON, William (ed.) *Aesthetics and Language* (Oxford, 1954).
23. FOOT, Philippa "Moral Beliefs," *Proceedings of the Aristotelian Society*, vol. 59 (1958-59).
24. FOSS, Martin *Symbol and Metaphor in Human Experience* (Princeton, N.J., 1942).
25. FREUD, Sigmund *Leonardo da Vinci* (New York, 1916).
26. FRIEDRICH, Carl A., BUKOFZER, Manfred F., HATZFELD, Helmut, MARTIN, John Rupert, STECHOW, Wolfgang. Assorted articles on the Baroque, *Journal of Aesthetics and Art Criticism*, vol. 14 (1955).
27. HAMPSHIRE, Stuart "Logic and Appreciation," in 22.
28. HARE, R. M. *The Language of Morals* (Oxford, 1952).
29. HART, H. L. A. "The Ascription of Responsibility and Rights," *Proceedings of the Aristotelian Society*, vol. 49 (1948-49).

30. HENLE, Paul "Metaphor," in Paul Henle (ed.), *Language, Thought and Culture* (Ann Arbor, Mich., 1958).
31. HENZE, Donald "Is the Work of Art a Construct?" *Journal of Philosophy*, vol. 52 (1955).
32. HEYL, Bernard C. "Meanings of Baroque," *Journal of Aesthetics and Art Criticism*, vol. 19 (1961).
33. — — *New Bearings in Aesthetics and Art Criticism* (New Haven, Conn., 1943).
34. — — "Relativism Again," *Journal of Aesthetics and Art Criticism*, vol. 5 (1946).
35. HOSPERS, John "The Concept of Artistic Expression" (slightly revised), in Morris Weitz (ed.), *Problems in Aesthetics* (New York, 1959).
36. — — "Implied Truths in Literature," *Journal of Aesthetics and Art Criticism*, vol. 19 (1960).
37. — — *Meaning and Truth in the Arts* (Chapel Hill, N.C., 1946).
38. — — "The Croce-Collingwood Theory of Art," *Philosophy*, vol. 31 (1956).
39. HUNGERLAND, Isabel C. "The Logic of Aesthetic Concepts," in *Proceedings and Addresses of the American Philosophical Association*, vol. 36 (1962-63) (Yellow Springs, Ohio, 1963).
40. — — *Poetic Discourse* (Berkeley, 1958).
41. — — "The Concept of Intention in Art Criticism," *Journal of Philosophy*, vol. 52 (1955).
42. — — "Contextual Implication," *Inquiry*, vol. 4 (1960).
43. ISENBERG, Arnold "Critical Communication," *Philosophical Review*, vol. 58 (1949).
44. — — "Perception, Meaning, and the Subject-matter of Art," *Journal of Philosophy*, vol. 41 (1944).
45. KENNICK, William E. "Does Traditional Aesthetics Rest on a Mistake?" *Mind*, vol. 67 (1958).
46. KNIGHT, Helen "The Use of 'Good' in Aesthetic Judgments," *Proceedings of the Aristotelian Society*, vol. 36 (1936).
47. KRIS, Ernst *Psychoanalytic Explorations in Art* (New York, 1952).
48. LANGER, Susanne *Feeling and Form* (New York, 1953).
49. — — *Philosophy in a New Key* (Cambridge, Mass., 1942).
50. — — *Problems of Art* (New York, 1957).
51. — — "On a New Definition of 'Symbol,'" in her *Philosophical Sketches* (Baltimore, 1962).
52. LANGEFELD, H. S. *The Aesthetic Attitude* (New York, 1920).
53. LEE, Vernon *The Beautiful* (Cambridge, 1913).
54. MACDONALD, Margaret "Some Distinctive Features of Arguments Used in Criticism in the Arts" (somewhat altered), in 22.
55. — — and SCRIVEN, Michael "The Language of Fiction" (symposium), *Proceedings of the Aristotelian Society, Supplementary Volume XXVII* (1954).
56. — — Review of 96, *Mind*, vol. 60 (1951).
57. — — "Art and Imagination," *Proceedings of the Aristotelian Society*, vol. 53 (1952-53).
58. MARITAIN, Jacques *Creative Intuition in Art and Poetry* (New York, 1953).
59. MARGOLIS, Joseph *The Language of Art and Art Criticism* (Detroit, 1965).
60. — — "The Logic of Interpretation," in Joseph Margolis (ed.), *Philosophy Looks at the Arts* (New York, 1962).
61. — — "Mr. Weitz and the Definition of Art," *Philosophical Studies*, vol. 9 (1958).
62. — — "Aesthetic Perception," *Journal of Aesthetics and Art Criticism*, vol. 19 (1960).
63. MORGAN, Douglas N. "Psychology and Art Today," *Journal of Aesthetics and Art Criticism*, vol. 9 (1950).
64. NAGEL, Ernest Review of 49, *Journal of Philosophy*, vol. 40 (1943).
65. NAHM, Milton C. *The Artist as Creator* (Baltimore, 1956).
66. NOWELL-SMITH, P. H. *Ethics* (London, 1954).
67. PASSMORE, J. A. "The Dreariness of Aesthetics," *Mind*, vol. 60 (1951).
68. PEPPER, Stephen *Aesthetic Quality* (New York, 1938).
69. — — *The Work of Art* (Bloomington, Ind., 1955).
70. — — "Further Considerations of the Aesthetic Work of Art," *Journal of Philosophy*, vol. 49 (1952).
71. PRALL, David *Aesthetic Analysis* (New York, 1936).
72. REDPATH, Theodore "Some Problems of Modern Aesthetics," in C. A. Mace (ed.), *British Philosophy in the Mid-Century* (London, 1957).
73. REID, L. A. *A Study in Aesthetics* (London, 1931).
74. RICHARDS, I. A. *The Philosophy of Rhetoric* (London, 1936).
75. — — *Science and Poetry* (revised, London, 1935).
76. SCHWYZER, H. R. G. "Sibley's 'Aesthetic Concepts,'" *Philosophical Review*, vol. 72 (1963).
77. SIBLEY, Frank "Aesthetic Concepts: A Rejoinder," *Philosophical Review*, vol. 72 (1963).
78. — — "Aesthetic Concepts" (with extensive minor corrections), in J. Margolis (ed.), *Philosophy Looks at the Arts* (New York, 1962).
79. — — "Aesthetics and the Looks of Things," *Journal of Philosophy*, vol. 56 (1959).
80. STERN, Gustaf "Meaning and Change of Meaning" in *Göteborgs Högskolas Årsskrift*, vol. 38: 1, 1932 (Göteborg, 1931).

81. STEVENSON, C. L. "On the Reasons That Can be Given for the Interpretation of a Poem," in J. Margolis (ed.), *Philosophy Looks at the Arts* (New York, 1962).
82. — — *Ethics and Language* (New Haven, Conn., 1944).
83. — — "Interpretation and Evaluation in Aesthetics," in Max Black (ed.), *Philosophical Analysis* (Ithaca, N.Y., 1950).
84. — — "Symbolism in the Non-representative Arts," in Paul Henle (ed.), *Language, Thought, and Culture* (Ann Arbor, Mich., 1958).
85. TOMAS, Vincent "The Concept of Expression in Art," (with minor corrections), in J. Margolis (ed.), *Philosophy Looks at the Arts* (New York, 1962).
86. — — (ed.), *Creativity in the Arts* (Englewood Cliffs, N.J., 1964).
87. — — "Aesthetic Vision," *Philosophical Review*, vol. 68 (1959).
88. URMSON, J. O. "What Makes a Situation Aesthetic?" *Proceedings of the Aristotelian Society*, vol. 31 (1957-58).
89. USHENKO, Andrew "Metaphor," *Thought*, vol. 30 (1955).
90. VIVAS, Eliseo *Creation and Discovery* (New York, 1955).
91. — — "Animadversions on Imitation and Expression," *Journal of Aesthetics and Art Criticism*, vol. 19 (1961).
92. — — "Contextualism Reconsidered," *Journal of Aesthetics and Art Criticism*, vol. 18 (1959).
93. WACKER, Jeanne "Particular Works of Art," *Mind*, vol. 69 (1960).
94. WEISS, Paul *The World of Art* (Carbondale, Ill., 1961).
95. WEITZ, Morris *Philosophy of the Arts* (Cambridge, Mass., 1950).
96. — — "The Role of Theory in Aesthetics," *Journal of Aesthetics and Art Criticism*, vol. 15 (1956).
97. WIMSATT, W. K., Jr. *The Verbal Icon* (Lexington, Ky., 1954).
98. ZIFF, Paul "Reasons in Art Criticism," in I. Scheffler (ed.), *Philosophy and Education* (Boston, 1958).
99. — — "Art and the 'Object of Art'," *Mind*, vol. 60 (1951).
100. ZINK, Sidney "The Poetic Organism," *Journal of Philosophy*, vol. 42 (1945).

SUPPLEMENTARY BIBLIOGRAPHY

- AIKEN, Henry "A Pluralistic Analysis of Aesthetic Value," *Philosophical Review*, vol. 59 (1950).
- — "The Aesthetic Relevance of Belief," *Journal of Aesthetics and Art Criticism*, vol. 9 (1951).
- — "Some Notes Concerning the Aesthetic and the Cognitive," *Journal of Aesthetics and Art Criticism*, vol. 13 (1955).
- AMES, Van Meter "John Dewey as Aesthetician," *Journal of Aesthetics and Art Criticism*, vol. 12 (1953).
- ASCHENBRENNER, Karl "Aesthetic Theory—Conflict and Conciliation," *Journal of Aesthetics*, vol. 18 (1959).
- — "Intention and Understanding," in *Meaning and Interpretation. University of California Publications in Philosophy*, vol. 25 (Berkeley, 1950).
- — "Critical Reasoning," *Journal of Philosophy*, vol. 57 (1960).
- — "Creative Receptivity," *Journal of Aesthetics and Art Criticism*, vol. 22 (1963).
- — and ISENBERG, Arnold (eds.) *Aesthetic Theories: Studies in the Philosophy of Art* (Englewood Cliffs, N.J., 1965).
- BEARDSLEY, Monroe "Representation and Presentation: A Reply to Professor Dickie," *Journal of Philosophy*, vol. 58 (1961).
- — MORGAN, Douglas N., and MOTHERSILL, Mary "Symposium: On Arts and the Definition of Arts," *Journal of Aesthetics and Art Criticism*, vol. 20 (1961).
- — "The Metaphorical Twist," *Philosophy and Phenomenological Research*, vol. 22 (1962).
- BOAS, George *Wingless Pegasus* (Baltimore, 1960).
- — "The Problem of Meaning in the Arts," in *Meaning and Interpretation. University of California Publications in Philosophy*, vol. 25 (Berkeley, 1950).
- BULLOUGH, Edward *Aesthetics* (London, 1957).
- CARVER, G. A. *Aesthetics and the Problem of Meaning* (New Haven, 1952).
- COHEN, Marshall "Appearance and the Aesthetic Attitude," *Journal of Philosophy*, vol. 56 (1959).
- CROCKETT, Campbell "Psychoanalysis in Art Criticism," *Journal of Aesthetics and Art Criticism*, vol. 17 (1958).
- DICKIE, George "Design and Subject Matter: Fusion and Confusion," *Journal of Philosophy*, vol. 58 (1961).
- DUCASSE, C. J. "The Sources of the Emotional Import of an Aesthetic Object," *Philosophy and Phenomenological Import*, vol. 21 (1961).
- EVELING, H. S. "Composition and Criticism," *Proceedings of the Aristotelian Society*, vol. 59 (1958-59).
- GALLIE, W. B. "Art as an Essentially Contested Concept," *Philosophical Quarterly*, vol. 6 (1956).
- GARVIN, Lucius "Emotivism, Expression, and Symbolic Meaning," *Journal of Philosophy*, vol. 55 (1958).

- GOTTSHALK, D. W. "Aesthetic Expression," *Journal of Aesthetics and Art Criticism*, vol. 13 (1954).
- HAMPSHIRE, Stuart *Feeling and Expression* (London, 1961).
- HANNAY, A. H., HOLLOWAY, John, and MACDONALD, Margaret, "What are the Distinctive Features of Arguments Used in Art Criticism?" (symposium), *Proceedings of the Aristotelian Society, Supplementary Volume XXIII* (1949).
- HARRE, R. "Quasi-Aesthetic Appraisals," *Philosophy*, vol. 33 (1958).
- HARRISON, Andrew "Poetic Ambiguity," *Analysis*, vol. 23 (1963).
- HARRISON, B. "Some Uses of 'Good' in Criticism," *Mind*, vol. 69 (1960).
- HENDERSON, G. P. "An 'Orthodox' Use of the Term 'Beautiful,'" *Philosophy*, vol. 35 (1960).
- HENZE, Donald "The Work of Art," *Journal of Philosophy*, vol. 54 (1957).
- "The 'Look' of a Work of Art," *Philosophical Quarterly*, vol. 2 (1961).
- HEPBURN, Ronald W. "Literary and Logical Analysis," *Philosophical Quarterly*, vol. 8 (1958).
- "Emotions and Emotional Qualities," *British Journal of Aesthetics*, vol. 1 (1961).
- HEYL, Bernard C. "Artistic Truth Reconsidered," *Journal of Aesthetics and Art Criticism*, vol. 8 (1950).
- HOFFMAN, Robert "Aesthetic Argument—Interpretive and Evaluative," *Philosophical Quarterly*, vol. 2 (1961).
- HOFSTADTER, Albert "Art and Spiritual Validity," *Journal of Aesthetics and Art Criticism*, vol. 22 (1963).
- and KUHN, Richard (eds.) *Philosophies of Art and Beauty* (New York, 1964).
- HOSPERS, John "Literature and Human Nature," *Journal of Aesthetics and Art Criticism*, vol. 17 (1958).
- HUNGERLAND, Helmut "The Aesthetic Response Re-considered," *Journal of Aesthetics and Art Criticism*, vol. 16 (1957).
- INGARDEN, Roman "Aesthetic Experience and Aesthetic Object," *Philosophy and Phenomenological Research*, vol. 12 (1961).
- ISENBERG, Arnold "The Esthetic Function of Language," *Journal of Philosophy*, vol. 46 (1949).
- *Analytic Philosophy and the Study of Art* (privately circulated, 1950).
- "The Problem of Belief," *Journal of Aesthetics and Art Criticism*, vol. 13 (1955).
- JENKINS, Iredell "The Being and the Meaning of Art," *Review of Metaphysics*, vol. 14 (1961).
- JESSUP, Bertram "Taste and Judgment in Aesthetic Experience," *Journal of Aesthetics and Art Criticism*, vol. 19 (1960).
- KADISH, Mortimer R. and HOFSTADTER, Albert "Symposium: The Evidence for Esthetic Judgment," *Journal of Philosophy*, vol. 54 (1957).
- KAPLAN, Abraham "Obscenity as an Esthetic Category," in Sidney Hook (ed.), *American Philosophers at Work* (New York, 1956).
- and KRIS, Ernst "Esthetic Ambiguity," *Philosophy and Phenomenological Research*, vol. 8 (1948).
- "Referential Meaning in the Arts," *Journal of Aesthetics and Art Criticism*, vol. 12 (1954).
- KENNICK, William E. "Art and the Ineffable," *Journal of Philosophy*, vol. 58 (1961).
- (ed.) *Art and Philosophy* (New York, 1964).
- KHATCHADOURIAN, Haig "Works of Art and Physical Reality," *Ratio*, vol. 2 (1960).
- "Art-names and Aesthetic Judgments," *Philosophy*, vol. 36 (1961).
- KRISTELLER, Oscar "The Modern System of the Arts: A Study in the History of Aesthetics," *Journal of the History of Ideas*, vol. 12 (1951); vol. 13 (1952).
- KUHN, Richard "Criticism and the Problem of Intention," *Journal of Philosophy*, vol. 57 (1960).
- "Art Structures," *Journal of Aesthetics and Art Criticism*, vol. 19 (1960).
- LANGER, Susanne K. (ed.) *Reflections on Art* (Baltimore, 1958).
- LEVICH, Marvin (ed.) *Aesthetics and the Philosophy of Criticism* (New York, 1963).
- MARGOLIS, Joseph "The Identity of a Work of Art," *Mind*, vol. 67 (1959).
- "Describing and Interpreting Works of Art," *Philosophy and Phenomenological Research*, vol. 22 (1961).
- "Rejoinder to W. D. L. Scobie on 'The Identity of a Work of Art,'" *Mind*, vol. 70 (1961).
- "Creativity, Expression, and Value Once Again," *Journal of Aesthetics and Art Criticism*, vol. 22 (1963).
- (ed.) *Philosophy Looks at the Arts* (New York, 1962).
- MAYO, Bernard "Art, Language, and Philosophy in Croce," *Philosophical Quarterly*, vol. 5 (1955).
- MORGAN, Douglas "The Concept of Expression in Art," in *Science, Language, and Human Rights* (Philadelphia, 1952).
- "Creativity Today," *Journal of Aesthetics and Art Criticism*, vol. 12 (1953).
- "Icon, Index and Symbol in the Visual Arts," *Philosophical Studies*, vol. 6 (1955).
- MOTHERSILL, Mary "'Unique' as an Aesthetic Predicate," *Journal of Philosophy*, vol. 57 (1961).
- "Critical Reasons," *Philosophical Quarterly*, vol. 2 (1961).
- NAHM, Milton C. "The Philosophy of Aesthetic Expression: The Crocean Hypothesis," *Journal of Aesthetics and Art Criticism*, vol. 13 (1955).
- OSBORNE, Harold *Theory of Beauty* (London, 1952).
- *Aesthetics and Criticism* (London, 1955).
- PEPPER, Stephen C., and POTTER, Karl "The Criterion of Relevancy in Aesthetics: A Discussion," *Journal of Aesthetics and Art Criticism*, vol. 16 (1957).

- POLE, David "Varieties of Aesthetic Experience," *Philosophy*, vol. 30 (1955).
- "Morality and the Assessment of Literature," *Philosophy*, no. 37 (1962).
- PRICE, Kingsley B. "Is There Artistic Truth?" *Journal of Philosophy*, vol. 46 (1949).
- PURSER, J. W. R. *Art and Truth* (Glasgow, 1957).
- QUINTON, A. M. and MEAGER, Ruby "Tragedy" (symposium), *Proceedings of the Aristotelian Society, Supplementary Volume XXXIV* (1960).
- RADER, Melvin (ed.) *A Modern Book of Esthetics* (New York, 1960, 3d ed.).
- RUDNER, Richard "The Ontological Status of the Esthetic Object," *Philosophy and Phenomenological Research* (1950).
- "On Semiotic Aesthetic," *Journal of Aesthetics and Art Criticism*, vol. 10 (1951).
- "Some Problems of Nonsemiotic Aesthetic Theories," *Journal of Aesthetics and Art Criticism*, vol. 15 (1957).
- SAW, Ruth "Sense and Nonsense in Aesthetics," *British Journal of Aesthetics*, vol. 1 (1961).
- "What Is a 'Work of Art'?" *Philosophy*, vol. 36 (1961).
- and OSBORNE, Harold "Aesthetics as a Branch of Philosophy," *British Journal of Aesthetics*, vol. 1 (1960).
- SCOBIE, W. D. L. "Margolis on 'The Identity of a Work of Art'," *Mind*, vol. 69 (1960).
- SESONSKE, Alexander "Truth in Art," *Journal of Philosophy*, vol. 53 (1956).
- SPARSHOTT, Francis "Mr. Ziff and the 'Artistic Illusion'," *Mind*, vol. 61 (1952).
- *The Structure of Aesthetics* (Toronto, 1963).
- STEVENSON, C. L. "On 'What Is a Poem'?" *Philosophical Review*, vol. 66 (1957).
- "On the 'Analysis' of a Work of Art," *Philosophical Review*, vol. 67 (1958).
- STOLNITZ, Jerome "Notes on Comedy and Tragedy," *Philosophy and Phenomenological Research*, vol. 16 (1955).
- "On Objective Relativity on Aesthetics," *Journal of Philosophy*, Vol. 57 (1960).
- *Aesthetics and Philosophy of Art Criticism* (Cambridge, Mass., 1960).
- "Some Questions Concerning Aesthetic Perception," *Philosophy and Phenomenological Research*, vol. 22 (1961).
- "On the Origins of 'Aesthetic Disinterestedness,'" *Journal of Aesthetics and Art Criticism*, vol. 20 (1961).
- "Beauty": History of an Idea," *Journal of the History of Ideas*, vol. 23 (1961).
- TEJERA, V. "The Nature of Aesthetics," *British Journal of Aesthetics*, vol. 1 (1961).
- TOMAS, Vincent "Ducasse on Art and its Appreciation," *Philosophy and Phenomenological Research*, vol. 13 (1952).
- "Creativity in Art," *Philosophical Review*, vol. 67 (1958).
- TSUGAWA, Albert "The Objectivity of Aesthetic Judgments," *Philosophical Review*, Vol. 70 (1961).
- USHENKO, Andrew P. *Dynamics of Art* (Bloomington, Ind., 1953).
- "Pictorial Movement," *British Journal of Aesthetics*, vol. 1 (1961).
- VIVAS, Eliseo and KRIEGER, Murray (eds.) *The Problems of Aesthetics* (New York, 1953).
- *The Artistic Transaction and Essays on the Theory of Literature* (Columbus, Ohio, 1963).
- WALSH, Dorothy "Critical Reasons," *Philosophical Review*, vol. 69 (1960).
- WEISS, Paul *Nine Basic Arts* (Carbondale, Ill., 1961).
- *Religion and Art* (Milwaukee, 1963).
- WEITZ, Morris "Symbolism and Art," *Review of Metaphysics*, vol. 7 (1964).
- (ed.) *Problems in Aesthetics* (New York, 1959).
- *Philosophy in Literature* (Detroit, 1963).
- WELSH, Paul "Discursive and Presentational Symbols," *Mind*, vol. 64 (1955).
- "On Explicating Metaphors," *Journal of Philosophy*, vol. 60 (1963).
- WHEELWRIGHT, Philip *The Burning Fountain* (Bloomington, Ind., 1954).
- ZERBY, Lewis K. "A Reconsideration of the Role of Theory in Aesthetics—A Reply to Morris Weitz," *Journal of Aesthetics and Art Criticism*, vol. 16 (1957).
- ZIFF, Paul "The Task of Defining a Work of Art," *Philosophical Review*, vol. 63 (1953).
- "On What a Painting Represents," *Journal of Philosophy*, vol. 57 (1960).
- ZINK, Sidney "Is the Music Really Sad?" *Journal of Aesthetics and Art Criticism*, vol. 19 (1960).

III. THE PROBLEM OF THEORETICAL TERMS

PETER ACHINSTEIN*

PHILOSOPHERS with quite different viewpoints have considered it important to distinguish two sorts of terms employed by scientists. While various labels have been suggested I shall use the expressions *theoretical* and *non-theoretical* to represent the intended distinction. Those who propose it provide examples of terms which fall into these respective categories, and, although there is by no means general agreement on all classifications, there is substantial accord on many examples. What follows is a list of some of the illustrations cited:

<i>theoretical terms</i>	
electric field	mass
electron	electrical resistance
atom	temperature
molecule	gene
wave function	virus
charge	ego
<i>non-theoretical terms</i>	
red	floats
warm	wood
left of	water
longer than	iron
hard	weight
volume	cell nucleus

Some philosophers base this distinction on a concept of "observability." Others appeal to a notion of "conceptual organization" or "theory dependence." My purpose here is to examine this distinction and suggest reasons for doubting the claim that there is some unique criterion or set of criteria which underlies it. Rather, I shall argue, the notions of "observability," "conceptual organization," and "theory dependence" introduced by these authors generate many distinctions which result in a number of different ways of classifying

terms on the above lists. Since the alleged distinction between theoretical and non-theoretical terms has played an important role in the philosophy of science, as well as epistemology, an examination of its basis may show the need for reformulating some rather persistent issues and indicate the sort of steps which might profitably be taken.

I

The proposal first to be considered is that the above classification rests upon a distinction between entities or properties which are observable and those which are not. Thus, according to Carnap, we have on the one hand "terms designating observable properties and relations," and on the other, "terms which may refer to unobservable events, unobservable aspects or features of events."¹ Carnap does not go on to explain what he means by "observable" and "unobservable" and presumably believes that his readers will understand in at least a general way the distinction intended. Unfortunately the situation is more complex than he seems willing to admit, since the terms "observable" and "unobservable" can be employed for the purpose of making a substantial number of different points.

Consider the case of visual observation. Just what it is that I can appropriately claim to have observed depends very importantly on the particular context in which the claim is made.² Suppose that while sitting by the roadside at night I am asked what I observe on the road ahead. I might, in one and the same situation, reply in a number of different ways; for example: a car, the front of a car, a pair of automobile headlights, two yellowish lights, etc. Or, when driving on a dirt road in the daytime I might, in one and the same situation, claim to be observing a car, a trail produced by a car, or just a cloud of dust. Two points deserve emphasis here. First, in each case what I will

* The author is indebted to the National Science Foundation for support of research.

¹ Rudolf Carnap, "The Methodological Character of Theoretical Concepts," in Herbert Feigl and Michael Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 1 (Minneapolis, 1956), pp. 38-76.

² This is emphasized by J. L. Austin in *Sense and Sensibilia* (Oxford, 1962), pp. 97ff., in a discussion in which he is concerned with criticizing the doctrine that verbs of perception have different senses.

actually say that I have observed depends upon a number of factors, such as the extent of my knowledge and training, how much I am prepared to maintain about the object under the circumstances, and the type of answer I suppose my questioner to be interested in. Second, in both examples I might claim to have observed a car ahead, though in the first the only visible parts of the car are its headlights, and in the second I see none of its parts at all, not even a speck in the distance which *is* the car. Nor must such claims necessarily be deemed imprecise, inaccurate, ambiguous, or in any way untoward. In the particular circumstances it may be perfectly clear just what I am claiming, though someone ignorant of the context might misconstrue my claim and expect me to know more than I do, such as the color, the shape, or even the make of the car.

Both of these points can be illustrated by reference to scientific contexts. Suppose that an experimental physicist, acquainted with the sorts of tracks left by various subatomic particles in cloud chambers, is asked what he is now observing in the chamber. He might reply in a number of ways, e.g., electrons passing through the chamber, tracks produced by electrons, strings of tiny water droplets which have condensed on gas ions, or just long thin lines. Similarly, to the question "What does one observe in a cathode-ray experiment?" the physicist might answer: Electrons striking the fluorescent zinc sulfide screen, light produced when molecules of zinc sulfide are bombarded, a bright spot, etc. In each case what the physicist actually claims to have observed will depend upon how much he knows and is prepared to maintain, the knowledge and training of the questioner, and the sort of answer he thinks appropriate under the circumstances. Furthermore, just as in the previous automobile examples, there are situations in which the physicist, concerned mainly with indicating the occurrence of certain events and with proper identification, will report observing various particles pass through the chamber—electrons if the tracks are long and thin, alpha particles if they are shorter and heavier. Whether he chooses to describe the situation in this way will depend upon the sort of factors noted above.

Analogous considerations hold for other terms on the "theoretical" list. Thus the physicist may

report having observed the electric field in the vicinity of a certain charge, or he may describe what he did as having observed the separation of leaves in an electroscope; he may report to be observing the rise in temperature of a given substance, or simply the increase in length of the column of mercury in the thermometer, etc. Accordingly, those who seek to compile a list of "observational" terms must not do so on the basis of an assumption to the effect that there exists some unique way of describing what is observed in a given situation. Nor can a classification of electrons, fields, temperature, etc., as *unobservable* be founded simply on a claim that one cannot ever report observing such items. Or, at least, if such a claim is made it will need to be expanded and defended in a manner not attempted by Carnap.

Suppose then, Carnap were to acknowledge that scientists often describe what they have observed in different ways, and that physicists do speak of observing such things as subatomic particles in cloud chambers and electric fields in the vicinity of charges, when the main concern is to report the occurrence of a certain event or the presence of a certain type of entity. Still, he might urge, contextual considerations of the sort mentioned will be irrelevant when we consider, strictly speaking, what is really observed in such cases. For what the physicist (really) observes is not the electron itself (but only its track, or a flash of light), just as in the automobile examples, what is (really) observed is not the car itself (but only its headlights, or a dust trail). And when the physicist detects the presence of an electric field, what he observes is not the field itself, but (say), only the separation of leaves in the electroscope. In general, it might be said, the distinction desired can be drawn on the basis of the claim that items on the "theoretical" list are not themselves (really) observable.³

I do not want to deny that electrons, fields, temperature, etc., can be described in this way. Indeed, such a description may be invoked when the physicist begins to explain just what claims he is making when he speaks of observation in each of these cases. However, as will presently be indicated, when an expression such as "not itself (really) observable" is employed it makes sense only with reference to a specific context of observation and

³ Cf. R. B. Braithwaite, "Models in the Empirical Sciences," in E. Nagel, P. Suppes, and A. Tarski (eds.), *Logic, Methodology, and Philosophy of Science* (Stanford, 1962), p. 227: "... in all interesting cases the initial hypotheses of the theory will contain concepts which are ... not themselves observable (call these theoretical concepts); examples are electrons, Schrödinger wave unctons, genes, ego-ideals."

some particular contrast. And I want to show that this fact precludes the general sort of distinction desired.

Suppose that in the second automobile example I report that I cannot (really) observe the car itself. What claim am I making? I might be saying that all I can observe is a dust trail and no speck in the distance which I can identify as the car; or that I can observe only a speck, but not the body of the car; or again, that I can see the car in the mirror but not with the naked eye. And obviously other contrasts could be cited which would make the point of my assertion clear.

Consider now the case of a virus examined by means of an electron microscope. Suppose I say that the microbiologist does not (really) observe the virus itself. What am I claiming? I might be saying that since (let us suppose) he employs a staining technique, he sees not the virus but only the staining material known to be present in certain parts of the specimen. Or, I might simply be saying that what he observes is the image of the object as presented by the microscope, and not the object itself. On the other hand, comparing electron microscopy with X-ray diffraction, I might claim that in the former case he is able to observe the virus itself, whereas in the latter case he observes only the effect of X rays on the virus. On different occasions any one of these contrasts, and others, could underlie the claim that the virus itself cannot (or can) really be observed.

Similar considerations are relevant in understanding a corresponding claim regarding electrons. In the context of a cloud chamber experiment, if I assert that electrons themselves cannot (really) be observed I might be saying that though a track is visible there is no speck which can be identified as the electron, in the way that if a jet airplane is close enough, one can see not only its trail but identify a certain speck in the distance as the airplane itself. On the other hand, I might wish to contrast the case of the electron with that of the neutron and claim that whereas electrons themselves *can* be observed in a cloud chamber, neutrons cannot. The point of this contrast is that neutrons, being neutral in charge, cannot cause ionization in their passage through the chamber, and hence will not produce a track, the way electrons will.⁴

In general, then, it is not sufficient simply to refer

to the previous list of so-called theoretical terms and claim that the distinguishing feature of the items designated by these terms is that they are not themselves (really) observable. What must be done is to indicate for each item the point of such a classification; and this is most readily accomplished by contrasting the sense in which it is said to be (really) unobservable itself with the sense in which something else is claimed to be observable. Now I do not deny that one could supply a context for each item on the list in which it would be appropriate to speak of that item in the manner proposed. The important point is simply that these contexts, and the sorts of contrasts they may involve, will in general be quite different and will yield different classifications. Here are a few contrasts some of which have already been noted, namely those between:

- (a) objects such as electrons and alpha particles which are detected by means of their tracks, and objects such as cars and airplanes which can be seen together with or apart from their tracks.
- (b) objects such as neutrons and neutrinos which do not leave tracks in a cloud chamber, and those such as electrons and alpha particles which do.
- (c) objects such as smaller molecules which it is necessary to stain in order to observe with the electron microscope, and larger objects for which staining is unnecessary.
- (d) objects such as individual atoms which are too small to scatter electrons appreciably and hence cannot be seen with the electron microscope, and larger objects such as certain molecules which have significant scattering power and hence can be observed.
- (e) objects requiring illumination by electron beams which then must be transformed into light via impact with a suitable screen, and objects visible with ordinary light.
- (f) objects which can only be observed by the production of images in microscopes, and those which can be seen with the naked eye.
- (g) properties such as electrical resistance whose magnitudes must be (or are generally) calculated on the basis of measuring a number of other quantities, and those properties for which this is usually not necessary.
- (h) properties such as temperature for which some instrumentation is usually required (for determining differences), and those such as color for which it is not.

⁴ Another contrast sometimes invoked in atomic physics has to do with objects such as electrons for which (in accordance with the Heisenberg uncertainty relationships) the product of the uncertainties in position and simultaneous velocity is much greater than that for objects of considerably larger mass such as atoms and molecules. A particle of the former sort may be classified as (itself) unobservable where it is this particular contrast which is intended.

- (i) objects such as electric fields which are not the sorts of things to which mass and volume are ascribed, and objects such as solids to which they are.

Each of these contrasts, as well as others which could readily be cited, *might* be used to generate some sort of observational distinction. Yet the same item would be classified differently depending upon the particular distinction invoked. Electrons would be unobservable under (a) and (d) but observable under (b); heavy molecules would be unobservable under (e) and (f) but observable under (c) and (d); temperature would be unobservable under (h) but observable under (g). Also, under certain contrasts some items cannot be classified at all—electrons under (c), (e), and (f); heavy molecules under (a) and (b); temperature under (a)–(f). Accordingly, if contrasts of the type cited above must be invoked to provide significance for the claim that a certain entity or property is itself (really) unobservable, then the sort of distinction required by Carnap seems difficult if not impossible to draw.

II

I want now to consider two qualifications which some authors place on the notion of observation. The first is that one should not speak simply of observability, but of *direct* observability. Hempel, e.g., writes:

In regard to an observational term it is possible, under suitable circumstances, to decide by means of direct observation whether the term does or does not apply to a given situation. . . . Theoretical terms, on the other hand, usually purport to refer to not directly observable entities and their characteristics.⁵

In offering this criterion, Hempel, like Carnap, mentions no special or technical sense which he attaches to the phrase "direct observation." Nor does he elaborate upon its meaning, except to cite a few examples.⁶ He admits that his characterization

does not offer a precise criterion and that there will be borderline cases. Yet the problem involved is more complex than Hempel seems willing to allow and, contrary to his suggestion, does not just turn on the question of drawing a more precise "dividing line." For the expression "(not) directly observable," like "(un)observable itself," is one whose use must be tied to a particular context and to some intended contrast. Thus, if the physicist claims that electrons cannot be observed directly, he may simply mean that instruments such as cloud chambers, cathode-ray tubes, or scintillation counters are necessary. Here direct observability has to do with observation by the unaided senses. Or, he might mean that when one observes an electron in a cloud chamber one sees only its track but not, e.g., a speck which one would identify as the electron itself. Again the nuclear physicist might claim that particles such as neutrons and neutrinos are not directly observable, meaning that such particles cannot themselves produce tracks in a cloud chamber, unlike electrons and alpha particles, which, under this contrast, would be deemed directly observable. Another type of situation in which the expression "direct observation" might be invoked involves a contrast between properties, such as electrical resistance, whose magnitudes must be calculated by first measuring other quantities, and those, such as length, or temperature, for which this is often not necessary.⁷

In short, many contrasts can be invoked by the notion of direct observation,⁸ and a given item will be classified in different ways depending upon the particular one intended. An appeal to direct observation by itself does little to advance the cause of generating a unique distinction, and when such an appeal is spelled out in individual cases various distinctions emerge.

The second qualification sometimes placed on observability concerns the *number* of observations necessary correctly to apply a term or expression. Thus, in "Testability and Meaning," Carnap writes:

⁵ Carl G. Hempel, "The Theoretician's Dilemma," in H. Feigl, M. Scriven, and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II (Minneapolis, 1958), pp. 37–98.

⁶ Observations of "readings of measuring instruments, changes in color or odor accompanying a chemical reaction, utterances made. . . ."

⁷ Indeed, it is for this very reason that in thermodynamics pressure, volume, and temperature are frequently called directly observable properties of a thermodynamic system, whereas other thermodynamic properties such as internal energy and entropy are not.

⁸ Somewhat simpler ones may of course be presupposed when the expression "directly observable" is employed in more everyday situations. For example, a contrast between an object such as the bank robber's coat which is readily accessible to view and his revolver which is hidden under it during the robbery; or between an item such as this side of the moon's surface which can readily be observed and the far side which from our vantage point is always hidden from view.

A predicate ' P ' of a language L is called *observable* for an organism (e.g., a person) N , if, for suitable argument, e.g., ' b ', N is able under suitable circumstances to come to a decision with the help of a few observations about the full sentence, say ' $P(b)$ ', i.e., to a confirmation of either ' $P(b)$ ' or ' $\neg P(b)$ ' of such high degree that he will either accept or reject ' $P(b)$ '.⁹

Carnap does not, however, explain his qualification in sufficient detail and leaves some important questions unanswered. Is the number of observations to mean the number of times the object must be observed (or, if an experiment is in question, the number of times the experiment needs to be repeated) before a property can definitely be ascribed to the object? Or does Carnap perhaps mean the number of different characteristics of the object which need to be observed? Again, he may be thinking of the amount of preliminary investigation necessary before a final observation can be made.¹⁰ Or perhaps all of these considerations are relevant.

Yet whether an observation or experiment will need to be repeated, or many different characteristics of the item in question examined, or considerable preliminary investigation undertaken, depends not only on the nature of the object or property under examination, but also on the particular circumstances of the investigator and his investigation. One relevant factor will be the type of instrument employed and how easily the scientist has learned to manipulate it. The physicist familiar with electroscopes need make few, if any, repetitions of an experiment with this instrument to determine the presence of an electric charge and hence of an electric field. Nor need he observe many characteristics of the field (e.g., its intensity and direction at a certain point) in order to determine its presence. And he will not always need to make extensive preliminary observations on the instrument but only a few. Yet, charges and electric fields are alleged to be unobservable. Another factor determining the facility with which an observer will identify an object or property is the extent of his knowledge regarding the particular

circumstances of the observation. If the physicist knows that a certain radioactive substance has been placed in a cloud chamber he may readily be able to identify the particles whose tracks are visible in the chamber. Whereas, if he knows nothing about the circumstances of the experiment, and he is simply shown a photograph of its results, successful identification may be a more complicated task. Rapid classification, then, depends in considerable measure upon particular features of the context of observation and the knowledge of the investigator. Under certain "suitable circumstances" (to use Carnap's phrase), quite a number of terms classified by him as non-observational can be correctly applied "with the help of [just] a few observations."

Furthermore, if non-observability in Carnap's sense is held to be sufficient for a *theoretical* classification additional difficulties emerge. For many fairly ordinary expressions are such that in numerous circumstances more than a "few observations" might well be required before correct application is possible; for example, "is chopped sirloin," "is a bridge which will collapse," "was composed by Corelli." Yet these do not really seem to be the sorts of expressions the authors in question wish to call theoretical. On the other hand, if, following Scheffler,¹¹ Carnap's criterion for "non-observability" is to be construed simply as a necessary but not a sufficient condition for being classified as theoretical, then unless further criteria are proposed (which they are not either by Carnap or Scheffler) we will have no general basis for separating theoretical from non-theoretical terms.¹² And if the above criterion concerning the number of observations is to be construed as a sufficient one for a non-theoretical classification (as Carnap and Scheffler suggest), then, as we have seen, many terms classified by these authors as theoretical will, in numerous situations, require reclassification.

III

I have considered attempts to base a theoretical-non-theoretical distinction upon some notion of

* Reprinted in Herbert Feigl and May Brodbeck (eds.), *Readings in the Philosophy of Science* (New York, 1953), pp. 47-92. Quotation from p. 63. Similar qualifications on observability have been expressed more recently by Grover Maxwell, "The Ontological Status of Theoretical Entities," in H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. III (Minneapolis, 1962), and by Israel Scheffler, *The Anatomy of Inquiry* (New York, 1963), p. 164.

¹⁰ According to Carnap when instruments are used we have "to make a great many preliminary observations in order to find out whether the things before us are instruments of the kind required." ("Testability and Meaning," p. 64.)

¹¹ *Op. cit.*, pp. 164ff.

¹² Scheffler concludes (*op. cit.*, p. 164) that the only thing left to do is simply specify an exhaustive list of primitive terms to be called "observational" (and presumably a corresponding list of those to be called "theoretical"). But this leaves the question of the basis for this separation unanswered.

observation. Generally speaking, the thesis that a list of "observational" terms can be compiled is defended by those envisaging the possibility of an "empiricist language." One of the underlying assumptions of this program appears to be that there exists a unique (or at least a most suitable) way of describing what is, or can be (really, directly) observed—a special "physical object" or "sense datum" vocabulary eminently fit for this task. Yet, as emphasized earlier, there are numerous ways to describe what one (really, directly) observes in a given situation, some more infused with concepts employed in various theories than others. This does not, of course, preclude the possibility of classifying certain reports as observational in a given case. The point is simply that there is no special class of terms which must be used in describing what is observed. Words from the previous "theoretical" list, such as "electron," "field," and "temperature," are frequently employed for this purpose.

Still, it might be urged, even though terms on both lists can be used in descriptions of what is (really, directly) observed, those on the first list are more "theory-dependent" than those on the second. And while it may not be possible to draw the intended large-scale distinction on the basis of observation, it is nevertheless feasible and important to separate terms on the basis of their "theoretical" character. It is to the latter position, which has been defended by Hanson and Ryle, that I now wish to turn.

According to Hanson a distinction should be drawn between terms which "carry a conceptual pattern with them," and terms which "are less rich in theory, and hence less able to serve in explanations of causes";¹³ or, in Ryle's words, between expressions which are "more or less heavy with the burthen of [a particular] theory . . . [and those which] carry none [of the luggage] from that theory."¹⁴ As an example of a term which carries with it a conceptual pattern Hanson cites the word "crater":

Galileo often studied the Moon. It is pitted with holes and discontinuities; but to say of these that they are craters—to say that the lunar surface is craterous—is to infuse theoretical astronomy into one's observations. . . . To speak of a concavity as a crater is to

commit oneself as to its origin, to say that its creation was quick, violent, explosive. . . .¹⁵

"Crater," then, carries with it a conceptual pattern not borne by (non-theoretical) terms such as "hole," "discontinuity," or "concavity."

Two notions underlie these suggestions, one stressed more by Hanson, the other by Ryle. The first is that a theoretical term is one whose application in a given situation can organize diffuse and seemingly unrelated aspects of that situation into a coherent, intelligible pattern; terms which carry no such organizing pattern Hanson sometimes calls "phenomenal." The second notion is that theoretical terms are such that "knowing their meanings requires some grasp of the theory" in which they occur.¹⁶ "The special terms of a science," Ryle asserts, "are more or less heavy with the burthen of the theory of that science. The technical terms of genetics are theory-laden, laden, that is, not just with theoretical luggage of some sort or other but with the luggage of genetic theory."¹⁷

Despite the fine examples Hanson cites, and the use he makes of the notion of "organizing patterns" in supplying trenchant criticisms of various philosophical positions, surely the first proposal fails to provide a sufficient characterization of those terms Hanson calls "theory-laden" (or "theory-loaded"). For almost any term can be employed in certain situations to produce the type of pattern envisaged. Indeed, Hanson himself offers many examples of this; quite early in his book, for instance, he presents a drawing whose meaning is incomprehensible until it is explained that it represents a bear climbing a tree. In this context, the expression "bear climbing a tree" is one which organizes the lines into an intelligible pattern. Moreover, one could describe contexts in which Hanson's "phenomenal" terms such as "hole," "concavity," "solaroid disc," etc., might be employed to organize certain initially puzzling data. Conversely, there are situations in which terms such as "crater," "wound," "volume," "charge," and "wave-length," which Hanson calls "theory-loaded," are used to describe data which are initially puzzling and require "conceptual organization." Hanson, at one point, grants that terms can have this dual function:

¹³ N. R. Hanson, *Patterns of Discovery* (Cambridge, England, 1958), p. 60.

¹⁴ Gilbert Ryle, *Dilemmas* (Cambridge, England, 1956), pp. 90-91.

¹⁵ *Op. cit.*, p. 56.

¹⁶ Ryle, *op. cit.*, p. 90.

¹⁷ *Ibid.*

It is not that certain words are absolutely theory-loaded, whilst others are absolutely sense-datum words. Which are the data words and which are the theory-words is a contextual question. Galileo's scar may at some times be a datum requiring explanation, but at other times it may be part of the explanation of his retirement.¹⁸

Yet this admission is a large one. For it means that the rendering of "conceptual organization" is not a special feature of terms on the "theoretical" list which sets them apart from those on the "non-theoretical" list. Whether a term provides an organizing pattern for the data depends on the particular situation in which it is employed. In some contexts, the use of the term "electron" will serve to organize the data (e.g., tracks in a cloud chamber); in others the term "electron" will be employed to describe certain data requiring organization (e.g., the discontinuous radiation produced by electrons in the atom). But even a presumably non-theoretical expression such as "(X is) writing a letter" can be used in certain contexts to organize data thereby explaining a piece of behavior, and in others to describe something which itself demands explanation.

Hanson does make reference to the "width" of terms, claiming that some expressions are "wider" theoretically than others and hence presumably "carry a [greater] conceptual pattern with them."¹⁹ So despite the fact that most terms are capable of serving explanatory functions, distinctions might still be drawn on the basis of "width." Yet it is not altogether clear how this metaphor should be unpacked. Sometimes the suggestion appears to be that one term will be "wider" than another if it can be used to explain situations whose descriptions contain the latter term. Thus, referring to the Galileo example quoted above, the term "scar" would be wider than the term "retirement" because it can be used to explain something designated by the latter. Yet this is not altogether satisfactory since we might imagine a case in which a man's retirement constituted part of an explanation of a scar he incurred. Again, the term "electron" can be employed in explanations of magnetic fields; yet the presence of a magnetic field can explain motions of electrons.

Perhaps, however, the reference to "width" should be understood in connection with the thesis, proposed by Ryle (and shared by Hanson), that certain terms depend for their meaning upon a particular theory, whereas others do not, or at least are much less dependent. By way of explanation Ryle cites as an analogy the situation in games of cards. To understand the expression "straight flush" one must know at least the rudiments of poker, whereas this is not so with the expression "Queen of Hearts," which is common to all card games and carries with it none of the special "luggage" of any of them. In a similar manner certain terms used by scientists are such that to understand them one must have at least some knowledge of the particular theory in which they appear. These are the "theory-laden" terms.²⁰ Other expressions utilized by scientists can be understood without recourse to any specific theoretical system.

Before examining this proposal some preliminary points should be noted. First, Ryle often seems to be suggesting that a theory-laden term is one which "carries the luggage" of one particular theory. Yet quite a few of the terms which he and Hanson classify as "theory-laden"—terms such as "temperature," "wave-length," "electron"—appear in, and might be thought to be infused with the concepts of, *many* scientific theories. Such terms are not restricted to just one theory, as with respect to games, "straight flush" is to poker. Second, it is certainly not a feature characteristic only of terms Ryle calls "theory-laden"—or only of such terms and those from card games—that they must be understood by reference to some scheme, system of beliefs, or set of facts. Following Ryle's lead one might draw up many different sorts of classifications, such as "university-laden" terms ("hour examination," "credit," and "tutorial") which cannot be understood without some knowledge of universities and their procedures; or, referring to scientific contexts, "instrument-laden" terms, such as "dial," "on," "off," which presumably would not appear on the "theoretical" list, yet require at least some knowledge of instruments and their uses. Thus Ryle's proposal must not be construed simply as a criterion for distinguishing terms which must

¹⁸ Hanson, *op. cit.*, pp. 59–60.

¹⁹ See *op. cit.*, p. 61.

²⁰ Cf. Hanson, *op. cit.*, pp. 61–62: "'Revoke,' 'trump,' 'finesse' belong to the parlance system of bridge. The entire conceptual pattern of the game is implicit in each term. . . . Likewise with 'pressure,' 'temperature,' 'volume,' 'conductor,' 'charge' . . . in physics. . . . To understand one of these ideas thoroughly is to understand the concept pattern of the discipline in which it figures."

be understood in the context of a set of beliefs from those which can be understood independently of any such set, though Hanson's "theory-laden" vs. "sense-datum" labels might misleadingly suggest this. Third, one must always specify the theory with respect to which a given term is or is not "theory-laden." And, it would appear, a term might receive this classification with reference to one theory but not another, though it occurs in both. For in one theory its meaning might not be understood unless the principles of the theory are known, whereas this would not necessarily be so in the case of the other theory (or at least there could be significant differences of degree). For example, "mass" might be considered "theory-laden" with respect to Newtonian mechanics but not with respect to the Bohr theory of the atom in which it also appears, since it can be understood independently of the latter. Thus, presumably, not every term occurring in a given theory will be "theory-laden," just as not every term found in standard formulations of the rules and principles of poker—such as "sequence," and "card"—will be "poker-laden," to use Ryle's expression. But if a theory must always be specified with respect to which a given term is deemed "theory-laden," and if a term can be classified in this way with reference to one theory and not to another in which it appears, then lists of the sort compiled at the beginning of this paper cannot be legitimately constructed. The most that could be done would be to cite particular theories and for each one compose such lists indicating which terms are to be considered theoretical and which not for that theory.²¹

Yet even this task may be deemed incapable of fulfillment once we examine Ryle's claim that expressions are theory-laden if they are such that "knowing their meanings requires some grasp of the theory." For there are many different ways in which terms might be said to be dependent upon principles of a given theory, and it is not altogether clear whether any or all of these should be classified as cases of "meaning-dependence." Since, we have already seen, the issue of whether a term is "theory-laden" must be considered always with reference to a specific theory, let us suppose that we

are given such a theory. Within it we could expect to encounter the following sorts of terms or expressions which might be considered theory-dependent in some sense:

(1) A term or expression whose *definition* cannot be stated without formulating some law or principle of the theory in question. For example, "Newton's gravitational constant" could only be defined by invoking the law of universal gravitation. The expression "Bohr atom," frequently employed in atomic physics, is defined as one satisfying the postulates of the Bohr theory. "Electrical resistance" is defined by reference to Ohm's law, etc.

(2) A term or expression whose definition can be stated without formulating laws of the theory, but whose use must be *justified* by invoking some of these laws. In electrostatics an "electrostatic unit charge" (esu) is defined as one which when placed in a vacuum one centimeter away from a like equal charge will repel it with a force of one dyne. Such a definition proves useful provided that the force with which like charges repel each other in a vacuum depends upon their distance, which it does according to Coulomb's law.

(3) A term whose definition can be stated without formulating laws of the theory, yet which denotes some more or less complex expression which appears in a formula whose *derivation* in the theory will not be understood unless certain laws of that theory are known. Quite often various expressions utilized in the theory will not be considered thoroughly understood unless one knows "where they come from," i.e., how certain formulas containing these expressions are derived from more fundamental principles of the theory. This is true, for example, of the term "enthalpy" in thermodynamics, which is defined as " $U + pV$," where U is the internal energy of a system, p its pressure, and V its volume. One standard method for introducing this term is by considering a process of constant pressure and applying the first law of thermodynamics, arriving at an expression containing " $U + pV$."

(4) A term or expression referring to something x which the theory is designed to describe and explain in certain ways, and for which the question

²¹ Ryle at many points does seem to be pressing for a distinction between terms appearing within the context (broadly speaking) of a given theory (or system) which are "theory-laden" and those which may also appear in the same general context but are not. This is evidenced by the sorts of examples he chooses ("light wave" vs. "blue"; "straight flush" vs. "Queen of Hearts"), and also by the questions he raises ("How are the special terms of Bridge or Poker [e.g., "trump"] logically related to the terms in which the observant child describes the cards that are shown to him [e.g., "hearts"]?"). So, relativizing the distinction to terms employed in the context of a particular theory is not completely foreign to Ryle's thought, though on some occasions he does suggest that he intends a broader distinction, as between terms "laden" with some theory or other and those dependent on no theory whatever.

"What is (an) x ?" could be answered, at least in part, by considering principles of that theory. Very often this question, rather than "What does the term ' x ' mean?" will be asked, and an answer given not by reference to some formal definition of ' x ' (if indeed one exists) but to principles of the theory which characterize features of x . For example, suppose one were to ask, "What are electrons?" Many sorts of replies could of course be given depending upon the knowledge and interests of the inquirer. Part of the answer might involve references to the Bohr or quantum theories which describe the various energy states of electrons within the atom; to the band theory of solids which uses quantum mechanical results in describing properties which electrons manifest in conductors; to the theory of the chemical bond which describes the sharing of electrons by atoms, etc. By characterizing various properties of electrons, theories such as these provide answers to the question "What are electrons?" and in this sense the term "electron" might be considered theory-dependent with respect to each.

(5) A term or expression whose *role* in the theory can only (or best) be appreciated by considering laws or principles in which it appears.²² In most theories the roles of constituent terms can be examined from several points of view. One might simply consider whether and how a given term is needed for the purpose of *formulating* some of the principles of the theory (for example, how ' h ' (Planck's constant) is used in the formulation of two fundamental postulates of the Bohr theory). Once a theory has been stated one might ask whether and how a certain term affords a *simplification* or *concise expression* of other principles (as how the term ' s ' (entropy) facilitates the formulation of an equation combining the first and second laws of thermodynamics); or how it is used in *proofs* of important theorems. From a wider viewpoint, the role of a term might also be studied by considering how principles in which it functions serve to *explain* various phenomena (as how "resonance potential" in the Bohr theory is used in the explanation of electron transitions to different energy levels). Conversely, one might consider

the manner in which principles of the theory are used to explain various phenomena which the term itself designates.²³ Thus, the role played by the expression "discrete spectral lines" in the Bohr theory might be specified by showing how the postulates of that theory serve to explain the sort of phenomenon referred to by this expression.

The five sorts of theory-dependence mentioned reflect various factors which may be relevant in understanding a given term and represent at least some of the ways in which expressions employed by a theory might be deemed "theory-laden" in Ryle's sense, i.e., dependent (at least in part) on the theory for their meaning. No doubt others could be listed and within those already mentioned further distinctions drawn. Yet each type of dependence cited, if employed to generate a classification of terms, might well yield different results. On the basis of the first sort of dependence noted we could distinguish (a) terms whose definitions are usually given by reference to laws of the theory in question, from (b) terms usually defined independently of such laws. (With respect to thermodynamics, "entropy" might be considered an expression of the former sort, whereas "enthalpy" would be placed in the latter category.) On the basis of the second dependence we could distinguish (a) terms (such as "electrostatic unit charge") whose definitions require justification by appeal to some of the principles of the theory, from (b) terms (such as "electrostatic force") which are not specifically defined in that theory, or whose definitions require no special defense by reference to principles of the theory. Referring to the third type of dependence we might distinguish between (a) expressions usually introduced by reference to formulas in the theory whose derivations will not be understood unless principles of that theory are known, and (b) expressions appropriated with unchanged definitions from other theories.²⁴ Consideration of the fourth type of dependence might prompt a distinction between (a) terms referring to entities and properties which the theory describes in such a way that the question "What is (an) x ?" could be answered by considering principles of that theory, and (b) terms occurring in that theory for

²² Cf. Ryle, "The Theory of Meaning," in Max Black (ed.), *The Importance of Language* (Englewood Cliffs, N.J., 1962), p. 161.

²³ Hanson, it might be noted, suggests at one point that terms referring to something explained by a given theory are dependent in meaning upon that theory. (Thus, he claims, Tycho and Kepler, because they had different theories about the movement of the sun, attached different meanings to "sun." *Op cit.*, p. 7.) This thesis is also defended by P. K. Feyerabend, "Explanation, Reduction, and Empiricism," in H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. III (Minneapolis, 1962).

²⁴ In thermodynamics "enthalpy" would fall into the first category; "pressure" into the second.

which this question would not usually be answered by reference to the theory itself but perhaps to others. With respect to the Bohr theory terms such as "electron" and "atomic nucleus" would fall into the first category, terms such as "velocity," "acceleration," and "mass" would not. In the case of the fifth dependence several classifications seem possible depending upon the particular type of role considered. Thus, if we group together those terms whose (principal) role is construed as that of simply enabling certain postulates to be formulated, and call these "theory-laden," in general we might expect to get a different classification from that obtained by grouping together terms introduced mainly for the purpose of simplifying certain formulations. In the Bohr theory Planck's constant ' h ' might be classified in the first manner, ' \hbar ' (h divided by 2π) in the second. And if we consider all of the different roles cited earlier and attempt to draw a distinction between (a) terms at least some of whose roles cannot be fully understood unless principles of the theory are known, and (b) terms for which this is not necessary, we will find hardly any distinction left to be drawn. For if the roles of a term or expression employed within a given theory are explained by showing how it enables certain principles to be formulated, or how it simplifies formulations, or how it functions in proofs of various theorems, or how it is employed in certain postulates for purposes of explanation, or how it is used to refer to something explained by the theoretical postulates, then obviously principles of the theory will need to be cited. Yet surely almost any term employed in a given theory can be shown to play at least one of these roles, and thus when all such terms are grouped together as "theory-laden" the class of non-theoretical terms all but vanishes.

In short, the various types of dependence noted earlier generate several distinctions between expressions in a given theory. In the case of some of these the class of expressions to be considered "theory-laden" will be quite small (e.g., those given explicit definitions by reference to laws of the theory); in other cases it might be larger (e.g., expressions designating entities or properties which are such that the question "What is (an) x ?" could be answered by reference to principles of the theory); and in still other cases it would include practically every term employed by the theory (expressions serving at least one role which cannot be fully understood without a knowledge of some

of the principles of the theory). Moreover, as should be evident from some of the examples cited, a term classified as "theory-laden" in one of these senses would not necessarily be so classified in another. For these reasons a criterion of theory-dependence of the sort proposed by Ryle and Hanson not only precludes the construction of theory-independent lists of the type given at the beginning of this paper, but, even with respect to a specific theory, can give rise to various distinctions under which terms may be classified differently. On the other hand if all the various senses in which a term might be dependent upon a given theory are lumped together and a term classified as "theory-laden" if it conforms to any of these, then such a label would become useless for distinguishing between terms in a given theory since it would be applicable to almost all of them.

Our conclusions here are relevant also for those seeking to draw the broader distinction between terms laden with the concepts and principles of some theory or other and terms dependent upon no theory whatever. Since there are many sorts of theory-dependence, various distinctions become possible. And should any one of the criteria outlined earlier be considered sufficient to render a term "theoretical," few if any terms will escape this classification.²⁵

IV

We have been considering the doctrine that expressions employed by scientists can be divided into two sets. On one view the principle of division rests upon observation; on the other, upon conceptual organization or theory-dependence. What has been shown is not that divisions are impossible to make, but rather that the proposed criteria are capable of generating distinctions of many different sorts, each tied, in most cases rather specifically, to a particular context of observation or to a particular theory. Questions such as "Does the term refer to something which can be observed?" and "Does its meaning depend upon some theory?" are too nebulous to provide illuminating classifications.

This means that certain problems raised by philosophers need serious rethinking. For example, authors of logical empiricist persuasion introduce the following issue: If theoretical terms in science do not refer to what is observable, how can they be said to have meaning? The type of answer given

²⁵ Especially if, following Hanson, a term which refers to something explained by a theory is to be considered "theory-laden." See n. 23.

consists in treating such terms as uninterpreted symbols which gain meaning in the context of a theory by being related via "correspondence rules" to observational terms. The problem is then to make this idea precise by providing a "criterion of meaningfulness" for theoretical terms.²⁶ Yet in the absence of some definite basis for drawing the intended large-scale distinction between theoretical and observational terms such problems cannot legitimately be raised, at least not in this form. Again Ryle, beginning with the question, "How is the World of Physics related to the Everyday World?" suggests the need to restate it in a clearer way as, in effect, "How are the 'theory-laden' concepts of physics related to others?" Yet unless his notion of theory-dependence is made precise and is shown to generate some rather definite distinction one will be in doubt about which terms he is referring to and what special characteristic he is attributing to them. This by no means precludes questions concerning the manner in which terms employed by scientists are tied to observation and to theories. Such questions, however, should not be raised about observation *in general*, or theories *in general*, but about particular sorts of observations and about terms in specific theories.

The concept *electron*, for instance, is tied to observation in a manner different from that of *temperature*, *field*, or *molecule*. And while there are some similarities there are also important differences, so that epithets such as "not directly observable" are bound to prove unhelpful. The philosopher of science genuinely concerned with the relation between theory and observation must

begin by examining individual cases. Taking a clue from earlier discussion, one illuminating procedure consists in invoking a series of contrasts. How, for example, is the observation of electrons similar to and also unlike the observation of high flying jets, large molecules, neutrons, chairs and tables? Each such contrast can be invoked to bring out some quite definite point concerning the relation between electrons and observation. This constitutes one important way of dealing with certain questions raised by logical empiricists, though obviously in a much more specific form than they envisage. Employing a series of contrasts we can also proceed to discover how given concepts are theory-dependent, thus turning our attention, though again in a specific way, to issues raised by Ryle and Hanson.

It is important to consider the manner in which entities studied by the physicist are tied to observation and to theory. Yet the philosopher of science must not at the outset assume that items on the "theoretical" list constructed earlier are related to observation (or theory) in the same way, or even in the same *general* way. Some items can be grouped together as similar in certain observational respects (for instance, electrons and neutrons as being too small to observe with electron microscopes), though in other such respects there will be important differences (e.g., ionizing effects). If categories are invoked to mark the similarities which do exist they will need to be a good deal more specific than the "theoretical" and "non-theoretical" classifications too often presupposed by epistemologists and philosophers of science.

The Johns Hopkins University

²⁶ See, e.g., Carnap, "The Methodological Character of Theoretical Concepts"; Hempel, *op. cit.*; William W. Rozeboom, "The Factual Content of Theoretical Concepts," in *Minnesota Studies in the Philosophy of Science*, vol. III, *op. cit.*

IV. "INVESTIGATIONS INTO LOGICAL DEDUCTION: II"

GERHARD GENTZEN*

SECTION IV.

SOME APPLICATIONS OF THE HAUPTSATZ

§ 1.

Applications of the Hauptsatz to Propositional Logic

1.1

A TRIVIAL consequence of the *Hauptsatz* is the already known consistency of classical (and intuitionist) *predicate logic* (cf., e.g., D. Hilbert and W. Ackermann, *Grundzüge der Theoretischen Logik* [Berlin, 1928, 1st edition], p. 65): the sequent \rightarrow (which is derivable from every contradictory sequent $\rightarrow A \& \neg A$, cf. 3.21) cannot be the lower sequent of any inference figure other than of a cut and is therefore not derivable.

1.2. *Solution of the decision problem for intuitionist propositional logic.*

On the basis of the *Hauptsatz* we can state a simple procedure for deciding of a formula of propositional logic—i.e., a formula without object variables—whether or not it is classically or intuitionistically valid. (For classical propositional logic a simple solution has actually been known for some time, cf., e.g., p. 11 of Hilbert-Ackermann).

First we prove the following *lemma*:

A sequent in whose antecedent one and the same formula does not occur more than three times as an *S*-formula, and in whose succedent, furthermore, one and the same formula occurs no more than three times as an *S*-formula, will be called a "reduced sequent." The following lemma now holds:

1.21. Every *LJ*- or *LK*-derivation whose endsequent is reduced, may be transformed into an *LJ*- or *LK*-derivation with the same endsequent,

in which all sequents are reduced (and in which no cuts occur if the original derivation did not contain any).

Proof of this lemma: If we eliminate from the antecedent of a sequent, in any places whatever (possibly none), all *S*-formulae occurring more than once, and if we do the same independently in the succedent, so that eventually these formulae occur only once, twice, or three times, we obtain a sequent that will be called a "reduction instance of that other sequent."

From a reduction instance of a sequent we may obviously derive all other reduction instances of the same sequent by means of thinnings, contractions, and interchanges such that in the course of these operations only reduced sequents occur.

After these preliminary remarks we now transform the *LJ*- or *LK*-derivation in hand in the following way:

All *basic sequents* as well as the *endsequent* are left intact; they are already all reduced sequents.

The *D*-sequents which belong to an *inference figure* are transformed into reduction instances of these sequents in a way about to be indicated. By virtue of our preliminary remarks it does not matter if a sequent belonging to two different *D*-inference figures is in each case replaced by a *different reduction instance*, since one sequent derives very simply from the other by thinnings, contractions, and interchanges so that eventually another flawless derivation results. (The same holds for a sequent which, while belonging to an inference figure, is also a basic sequent or an endsequent, since it is of course a reduction instance of itself.)

The transformations of the inference figures are now carried out in the following way:

If a formula occurs more than once within Γ , it is eliminated from Γ , both from the upper sequents and the lower sequent, as many times (from the

* Originally published in the *Mathematische Zeitschrift*, vol. 39 (1935), pp. 405-431. The first part of these investigations appeared under the same title, *ibid.*, pp. 176-221, and was published in the *American Philosophical Quarterly*, vol. 1 (1964) pp. 288-306. The present English translation is by Mr. M. E. Szabo (Sir George Williams University, Montreal) whose thanks are due to his late wife Ann for her continued encouragement, and to Mr. Michael Dummett of All Souls College, Oxford, for reading the translation and suggesting improvements.

appropriate places) as is necessary to ensure that finally it occurs in Γ no more than once. The same procedure is used for Δ , Θ , and A (i.e., those series of formulae that are designated by these letters in the schema III, 1.21 and 1.22, of the inference figure under consideration.)

Having carried out the transformations described, we have now a derivation consisting only of reduced sequents. (An interchange where **D** equals **E** may form an exception, yet this figure would be an identical inference figure and could have been avoided.)

The lemma is thus proved.

Given the *Hauptsatz*, together with corollary III, 2.513, and the preceding lemma (1.21), it now holds that:

1.22. For every correctly reduced sequent, both intuitionist and classical, there exists an *LJ*- or *LK*-derivation without cuts consisting only of reduced sequents, and whose *D*-*S*-formulae are subformulae of the *S*-formula of that sequent.

1.23. Consider now a sequent not containing an object variable. We wish to decide whether or not it is intuitionistically or classically correct. We can begin by taking in its place an equivalent reduced sequent **Sq**.

The number of all reduced sequents whose *S*-formulae are subformulae of the *S*-formulae of **Sq** is obviously finite. The decision procedure may therefore be carried out without further complications in the following way:

We consider the finite system of sequents in question and investigate first of all, which of these sequents are basic sequents. Then we examine each of the remaining sequents to determine whether there occurs an inference figure in which the sequent in question is the lower sequent and in which there occur as upper sequents one or two of those sequents that have already been found to be derivable. If this is the case, the sequent is added to the derivable sequents. (All this is obviously decidable.) We continue in this way until either the sequent **Sq** itself turns out to be derivable, or until the procedure yields no other derivable sequents. In the latter case the sequent **Sq** (by virtue of 1.22) is not derivable at all in the calculus under consideration (*LJ* or *LK*). We have therefore succeeded in establishing the correctness of that sequent.

1.3. A new proof of the *nonderivability of the law of the excluded middle in intuitionist logic*.

Our decision procedure could have been

¹ Cf. n. 2 of Part I.

formulated in a way better suited to the needs of practical application; yet the above presentation (1.2) was intended only to indicate a possibility in principle.

As an example, we shall prove the nonderivability of the law of the excluded middle in intuitionist logic by a method independent of the decision procedure described (although this procedure would have to yield the same result.)

(This nonderivability has already been proved by Heyting¹ in a completely different way.)

The sequent in question is of the form $\rightarrow A \vee \neg A$. Suppose there existed an *LJ*-derivation for it. According to the *Hauptsatz* there then exists such a derivation without cuts. Its lowest inference figure must be a \vee -IS, for in all other *LJ*-inference figures either the antecedent of the lower sequent is not empty, or a formula occurs in the succedent whose terminal symbol is not \vee ; there might still be the case of a thinning in the succedent, but the upper sequent would then be a \rightarrow , which, by virtue of 1.1, is not derivable.

Hence either $\rightarrow A$ or $\rightarrow \neg A$ would have to be already derivable (without cuts).

(From the same considerations, incidentally, it follows in general: If $A \vee B$ is an intuitionistically correct formula, then either **A** or **B** is an intuitionistically correct formula. In classical logic this does not hold, as the example of $A \vee \neg A$ already shows.)

Now $\rightarrow A$ cannot be the lower sequent of any *LJ*-inference figure whatever (if it is not a cut), unless that figure is another thinning with \rightarrow for its upper sequent. Furthermore, since $\rightarrow A$ is not a basic sequent, it is thus not derivable.

The same considerations show that $\rightarrow \neg A$ is derivable only from $A \rightarrow$ by a \neg -IS figure, and $A \rightarrow$ is in turn derivable only from A , $A \rightarrow$, since A contains no terminal symbol. Continuing in this way, we always reach only sequents of the type A , A , \dots , $A \rightarrow$, but never a basic sequent.

Hence $A \vee \neg A$ is not derivable in intuitionist predicate logic.

§2.

A Strengthened Form of the Hauptsatz for Classical Predicate Logic

2.1. We are here concerned with the following *strengthening of the Hauptsatz*:

Suppose that we have an *LK*-derivation whose endsequent is of the following kind:

Each *S*-formula of this sequent contains \forall and

\exists -symbols at most at the beginning, and their scope extends over the whole of the remaining formula.

In that case, the given derivation may be transformed into an *LK*-derivation with the same endsequent and having the following properties:

1. It contains no cuts.
2. It contains a *D*-sequent, let us call it a "middle sequent," which is such that its derivation (and hence also the sequent itself) contains no \forall and \exists -symbols, and where no inference figures other than \forall -IS, \forall -IA, \exists -IS, \exists -IA occur in the rest of the derivation, the middle sequent included; nor do there occur any other structural inference figures.

2.11. The middle sequent divides the derivation, as it were, into an upper part belonging to propositional logic, and a lower part containing only \forall and \exists -introductions.

Concerning the form of the transformed derivation, the following may still be readily concluded: The lower part, from the middle sequent to the endsequent, belongs to only one branch, since only inference figures with one upper sequent occur in it. The *S*-formulae of the middle sequent are of the following kind:

Every *S*-formula in the antecedent of the middle sequent results from an *S*-formula in the antecedent of the endsequent by the elimination of the \forall and \exists -symbols (as well as the bound object variables beside them), and by the replacement of the bound object variables in the rest of the formula by certain free object variables. The same procedure is followed in the case of succedents.

This follows from the same consideration as in III, 2.512.

2.2. Proof of the Theorem (2.1).²

The transformation of the derivation is carried out in several steps.

2.21. We begin by applying the *Hauptsatz* (III, 2.5): The derivation may accordingly be transformed into a derivation without cuts.

2.22. Transformation of basic sequents containing a \forall or \exists -symbol:

By virtue of the properties of subformulae III, 2.513, such sequents can only have the form $\forall \mathbf{x} \mathbf{F} \mathbf{x} \rightarrow \forall \mathbf{x} \mathbf{F} \mathbf{x}$ or $\exists \mathbf{x} \mathbf{F} \mathbf{x} \rightarrow \exists \mathbf{x} \mathbf{F} \mathbf{x}$. They are transformed into (suppose *a* to be a free object variable not yet occurring in the derivation):

$$\frac{\mathbf{Fa} \rightarrow \mathbf{Fa}}{\forall \mathbf{x} \mathbf{F} \mathbf{x} \rightarrow \mathbf{Fa}} \forall\text{-IA} \quad \text{or} \quad \frac{\mathbf{Fa} \rightarrow \mathbf{Fa}}{\mathbf{Fa} \rightarrow \exists \mathbf{x} \mathbf{F} \mathbf{x}} \exists\text{-IS}$$

$$\frac{}{\forall \mathbf{x} \mathbf{F} \mathbf{x} \rightarrow \forall \mathbf{x} \mathbf{F} \mathbf{x}} \forall\text{-IS} \quad \text{or} \quad \frac{}{\exists \mathbf{x} \mathbf{F} \mathbf{x} \rightarrow \exists \mathbf{x} \mathbf{F} \mathbf{x}} \exists\text{-IA}$$

By repeating this procedure sufficiently often we can obviously eliminate all \forall and \exists -symbols from every basic sequent of the derivation.

2.23. We now perform a complete induction according to the "order" of the derivation, which is defined as follows:

Of the operational inference figures we call those belonging to the symbols $\&$, \vee , \neg , and \supset "propositional inference figures," and the rest, i.e., \forall -IS, \forall -IA, \exists -IS, \exists -IA, "predicate inference figures." To each predicate inference figure in the derivation we assign the following ordinal number:

We consider that branch of the derivation that extends from the lower sequent of the inference figure up to the endsequent of the derivation (including it) and count the number of lower sequents of the propositional inference figures occurring in it. Their number is the ordinal number.

The sum of the ordinal numbers of all predicate inference figures in the derivation is the *order of the derivation*.

We intend to reduce that order step by step until it becomes zero.

Note that once this has been achieved the rest of the proof of the theorem (2.1) is easily carried out: (The steps involved (2.232) will be such as to preserve the properties that were established in 2.21 and 2.22.)

2.231. In order to do so we assume that the derivation has already been reduced to order zero. From the endsequent we now proceed to the upper

² The following special case of theorem 2.1 has already been proved by Herbrand in a completely different way: If a formula *P*, in which the symbols \forall and \exists occur only at the beginning and span the whole formula (as we know, for every formula there exists a classically equivalent formula of this kind, cf. for example H.-A. p. 63) is classically derivable, then there is a sequent (the above middle sequent) whose antecedent is empty and each one of whose succedent formulae results from *P* by the elimination of the \forall and \exists symbols (including the variables next to them) and by the replacement of the bound object variables by free variables. Furthermore, this sequent is classically derivable without the use of \forall and \exists -symbols, and from it we can derive $\rightarrow P$ simply by using those of our deduction figures which we have designated by \forall -IS, \exists -IS, contraction in the succedent and interchange in the succedent. (In addition to theorem 2.1, we would still have to consider the case of a thinning in the succedent, but it is easy to see that such instances can always be avoided.)

Cf. also: J. Herbrand, "Sur le problème fondamental de la logique mathématique," *Comptes rendus de la Société des sciences et des lettres de Varsovie*, (Classe III), vol. 24 (1931), p. 31, n. 1.

sequent of the inference figure above it. We stop as soon as we encounter the lower sequent of a propositional inference figure or a basic sequent; that sequent we call **Sq**. (It will serve us as "middle sequent," once it has been transformed in a way about to be indicated.)

The derivation of **Sq** is now transformed as follows:

We simply omit all *D-S*-formulae which still contain the symbols \forall and \exists . The above derivation remains correct after the described operation since, by virtue of 2.22, its basic sequents are not affected. Furthermore, no principal or side formula of an inference figure has been eliminated, for if such a formula had contained a symbol \forall or \exists , the principal formula would certainly have contained that symbol. Yet no predicate inference figures occur (if they did, the ordinal number of the inference figure would be greater than zero), and by virtue of the properties of subformulae (III, 2.513) and the hypothesis of theorem 2.1, the principal formulae of the propositional inference figures cannot contain a \forall or \exists . Now every inference figure remains correct if we eliminate, wherever it occurs as an *S*-formula in the figure, a formula which occurs neither as a principal nor as a side formula. This is easily seen from the schemata III, 1.21 and III, 1.22. (At worst, an identical inference figure may result, which is then eliminated in the usual manner.)

The sequent **Sq***, which has resulted from **Sq** by this transformation, differs from **Sq** in that certain *S*-formulae may possibly have been eliminated. We let the sequent **Sq*** be followed by several thinnings and interchanges such that in the end the sequent **Sq** reappears, and to it we attach the unaltered lower part of the derivation.

We have now reached our goal: **Sq*** is the "middle sequent," and it obviously satisfies all conditions imposed on the latter by theorem (2.1).

2.232. It now remains for us to carry out the *induction step* of our proof, i.e., the order of the derivation is assumed to be greater than zero, and our task is to diminish it.

2.232.1. We begin by *redesignating the free object variables* in the same way as in III, 3.10. As a result of this, the derivation has the following property (III, 3.101):

For every \forall -*IS* (or \exists -*IA*) it holds that the proper variable in the derivation occurs only in the sequent above the lower sequent of the \forall -*IS* (or \exists -*IA*) and does furthermore not occur in any other \forall -*IS* or \exists -*IA* as a proper variable.

The order of the derivation is hereby obviously left unchanged.

2.232.2. We now come to the transformation proper.

To begin with, we observe that in the derivation there occurs a predicate inference figure—let us call it **If1**—with the following property: If we follow that branch of the inference figure which extends from the lower sequent to the endsequent, then the first lower sequent of an operational inference figure reached is the lower sequent of a *propositional* inference figure (that inference figure we call **If2**). If there were no such instance, the order of the derivation would be equal to zero.

Now our aim is to slide the inference figure **If1** lower down in the derivation beyond **If2**. This is easily done by means of the following schemata:

2.232.21. Suppose that **If2** has *one* upper sequent.

2.232.211. Suppose that **If1** is a \forall -*IS*. Then that part of the derivation on which the operation is to be carried out runs as follows:

$$\frac{\frac{\Gamma \rightarrow \Theta, Fa}{\Gamma \rightarrow \Theta, \forall x Fx} \quad \forall -IS}{\Delta \rightarrow A} \quad \text{If2, possibly preceded by structural inference figures}$$

This we transform into:

$$\frac{\frac{\frac{\Gamma \rightarrow \Theta, Fa}{\Gamma \rightarrow Fa, \Theta, \forall x Fx} \quad \text{possibly several interchanges, as well as a thinning}}{\Delta \rightarrow Fa, A} \quad \text{possibly several interchanges}} \quad \left\{ \begin{array}{l} \text{inference figures of exactly the} \\ \text{same kind as above, i.e., If2,} \\ \text{possibly preceded by structural} \\ \text{inference figures} \end{array} \right.$$

$$\frac{\frac{\Delta \rightarrow A, Fa}{\Delta \rightarrow A, \forall x Fx} \quad \forall -IS}{\Delta \rightarrow A} \quad \text{possibly several interchanges and contractions}$$

The elimination of $\forall x Fx$ by contraction in the last step of the transformation is made possible by the fact that in A , $\forall x Fx$ must occur as an *S*-formula. (For the *S*-formula $\forall x Fx$ could not, in the original derivation, have been eliminated from the succedent by means of **If2** and the preceding structural inference figures, since it can obviously not be a side formula of **If2**, by virtue of the side-formula property III, 2.513 and the hypothesis of theorem 2.1.)

The restriction on variables is satisfied by the above \forall -*IS* (**If1**) by virtue of 2.232.1.

The order of the derivation has obviously been diminished by 1.

2.232.212. The case where **If**₁ is a \exists -IS is dealt with analogously; all we need do is to replace \forall by \exists .

2.232.213. The cases where **If**₁ is a \forall -IA or \exists -IA are treated as mirror images of the two preceding cases.

2.232.22. The case where **If**₂ has two upper sequents, i.e., $\&$ -IS, \vee -IA, or \supset -IA, can be dealt with quite correspondingly. At most a number of additional structural inference figures may be required.

2.3. Analogously to theorem 2.1 there are several ways in which the *Hauptsatz* may be further strengthened in the sense that certain restrictions can be placed on the order of occurrence of the operational inference figures in a derivation. For we can rearrange the inference figures to a large extent by sliding them above and beyond each other as was done above (2.232.2).

We shall not pursue this question further.

§3.

Application of the Strengthened Hauptsatz (2.1) to a New³ Consistency Proof for Arithmetic Without Complete Induction

By *arithmetic* we mean the (elementary, i.e., employing no analytical techniques) theory of the natural numbers. Arithmetic may be formalized by means of our logical calculus *LK* in the following way:

3.1. In arithmetic it is customary to employ "functions," e.g., x' (equals $x + 1$), $x + y$, $x \cdot y$. Yet since we did not introduce any functional symbols into our logical investigations, we shall, in order to make our logic nonetheless applicable to arithmetic, formalize the propositions of arithmetic in such a way that predicates take the place of functions. In place of the function x' , for example, we shall use the predicate $x\text{P}y$, which reads: x is the predecessor of y , i.e., $y = x + 1$. Furthermore, $[x + y = z]$ will be considered a predicate with three argument places. Thus the symbols $+$ and $=$ have here no independent meaning. A different predicate is $x = y$; the equality symbol here has thus no formal connection at all with the equality symbol in the previous predicate.

³ Earlier proofs may be found in the writings of J. von Neumann, "On Hilbert's Proof Theory," *Mathematische Zeitschrift*, vol. 26 (1927), pp. 1-46; J. Herbrand, "Sur la non-contradiction de l'Arithmétique," *Journal für die reine und angewandte Mathematik*, vol. 166 (1932), pp. 1-8.

The number 1, furthermore, will not be written as a symbol for an individual object, since we have only object *variables* in our logical formalism and no symbols for particular objects. We shall avoid this difficulty by saying that the predicate "One x " has the same intuitive meaning as " x is the number 1."

The sentence " $x + 1$ is the successor of x ," for example, could be rendered thus in our formalism: $\forall x \forall y \forall z ((\text{One } y \ \& \ [x + y = z]) \supset x\text{P}z)$.

All other natural numbers can be represented by the predicates $\text{One } x \ \& \ x\text{P}y$; $\text{One } x \ \& \ x\text{P}y \ \& \ y\text{P}z$, etc.

Yet how are we now to integrate into our calculus the *predicate symbols* just introduced, having admitted nothing but propositional variables? To do so we simply stipulate that the predicate symbols are to be treated in exactly the same way as propositional variables. More precisely: We regard expressions of the form

$\text{One } x, x\text{P}y, x = y, [x + y = z]$,

where any object variables stand for x, y, z , merely as more easily intelligible ways of writing the formulae

$Ax, Bxy, Cxy, Dxyz$.

In this sense the axiom formulae that follow are indeed *formulae* in accordance with our definition.

(Yet we cannot, of course, regard the number 1 as a way of writing an object variable, since in our calculus the object variables really function as *variables*, which is not so in the case of propositional variables.)

As "*axiom formulae*" of our *arithmetic* we shall initially take the following, and shall later, once the consistency proof has been carried out (cf. 3.3), state general criteria for the formation of further admissible axiom formulae:

Equality:	$\forall x (x = x)$	(reflexivity)
	$\forall x \forall y (x = y \supset y = x)$	(symmetry)
	$\forall x \forall y \forall z ((x = y \ \& \ y = z) \supset x = z)$	(transitivity)
One:	$\exists x (\text{One } x)$	(existence of 1)
	$\forall x \forall y ((\text{One } x \ \& \ \text{One } y) \supset x = y)$	(uniqueness of 1)
Predecessor:	$\forall x \exists y (x\text{P}y)$	(existence of successor)
	$\forall x \forall y (x\text{P}y \supset \neg \text{One } y)$	(1 has no predecessor)
	$\forall x \forall y \forall z \forall u ((x\text{P}y \ \& \ z\text{P}u \ \& \ x = z) \supset y = u)$	(uniqueness of successor)

$$\begin{array}{l} \forall x \forall y \forall z \forall u ((xPry \& zPr u \& y = u) \\ \supset x = u) \\ \text{(uniqueness of predecessor)} \end{array}$$

A formula **B** is called *derivable* in arithmetic without complete induction, if there is an *LK*-derivation for a sequent

$$A_1, \dots, A_\mu \rightarrow B$$

in which A_1, \dots, A_μ are axiom formulae of arithmetic.

The fact that this formal system does actually allow us to represent the types of proof customary in intuitive arithmetic (as long as they do not use complete induction) cannot be *proved*, since for considerations of an intuitive character no precisely demarcated framework exists. We can merely verify this in the case of individual intuitive proofs by attempting them.

3.2. We shall now prove the *consistency* of the formal system just presented. With the help of the strengthened *Hauptsatz* (2.1) our task is in fact quite simple.

3.21. A "contradiction formula" $A \& \neg A$ is derivable in our system if and only if there exists an *LK*-derivation for a sequent with an *empty succedent* and with arithmetic axiom formulae in the antecedent, viz.:

From $\Gamma \rightarrow A \& \neg A$ we obtain $\Gamma \rightarrow$ in the following way:

$$\begin{array}{c} \frac{A \rightarrow A}{\neg A, A \rightarrow} \neg\text{-}IA \\ \frac{\neg A, A \rightarrow}{A \& \neg A, A \rightarrow} \&\text{-}IA \\ \frac{A \& \neg A, A \rightarrow}{A, A \& \neg A \rightarrow} \text{interchange} \\ \frac{A, A \& \neg A \rightarrow}{A \& \neg A, A \& \neg A \rightarrow} \&\text{-}IA \\ \frac{\Gamma \rightarrow A \& \neg A \quad A \& \neg A \rightarrow}{\Gamma \rightarrow} \text{contraction cut} \end{array}$$

The converse is obtained by carrying out a thinning in the succedent.

Thus, if our arithmetic is inconsistent, there exists an *LK*-derivation with the endsequent

$$A_1, \dots, A_\mu \rightarrow,$$

where A_1, \dots, A_μ are arithmetic axiom formulae.

3.22. We now apply the strengthened *Hauptsatz* (2.1). The arithmetic axiom formulae fulfil the requirement laid down for the *S*-formulae of the endsequent. Hence there exists an *LK*-derivation

with the same endsequent which has the following properties:

1. It contains no cuts.

2. It contains a *D*-sequent, the "middle sequent," whose derivation contains no \forall and \exists -symbols, and whose endsequent results from a number of inference figures \forall -*IA*, \exists -*IA*, thinnings, contractions and interchanges in the antecedent. The middle sequent has an empty succedent (2.11).

3.23. We then proceed to redesignate the free object variables as in III, 3.10. All mentioned properties remain unchanged, and the following property is added (III, 3.101): The proper variable of each \exists -*IA* in the derivation occurs only in sequents *above* the lower sequent of the \exists -*IA*.

3.24. Then we replace every occurrence of a free object variable by one and the same natural number in a way to be described presently. In doing so we are left with a figure which we can no longer call an *LK*-derivation. We shall see later to what extent it is nevertheless intuitively meaningful.

The replacement of the free object variables by numbers is carried out in the following order:

3.241. First we replace all free object variables which do not occur as the proper variable of a \exists -*IA* by the number 1 throughout. (We could also take another number.)

3.242. Then we take every \exists -*IA* inference figure in the derivation, beginning with the lowest and taking each figure in turn, and replace each proper variable (wherever it occurs in the "derivation") by a number. That number is determined as follows:

The \exists -*IA* can only run:

$$\frac{\text{One } a, \Gamma \rightarrow \Theta}{\exists x \text{ One } x, \Gamma \rightarrow \Theta} \quad \text{or} \quad \frac{\nu \text{Pra}, \Gamma \rightarrow \Theta}{\exists y \nu \text{Pry}, \Gamma \rightarrow \Theta}$$

(by virtue of the subformula property III, 2.513; ν can be only one number, by virtue of 3.241 and 3.23). In the first case we replace a by 1, in the second case by the number that is one greater than ν .

3.25. Now we examine the figure which has resulted from the derivation. We are particularly interested in what the (former) middle sequent now looks like. We can say this about it:

Its succedent is empty, and each of the antecedent *S*-formulae either has the form $\text{One } 1$ or νPrv , where a number stands for ν , and where a number one greater than the previous one stands for ν' ; or it results from an arithmetic axiom form-

ula that has only \forall -symbols at its beginning, by the elimination of the \forall -symbols (and the bound object variables next to them) and the substitution of *numbers* for the bound object variables in the remaining part of the formula. (All this follows from the same consideration as in III, 2.512, also cf. 2.11.)

Thus, the *S*-formulae in the antecedent of the middle sequent represent *intuitively correct numerical propositions*. It further holds for the "derivation" of the middle sequent that it has resulted from a derivation containing no \forall or \exists -symbols, by having all its occurrences of free object variables replaced by numbers. Intuitively, such a "derivation" constitutes in effect a *proof in arithmetic using only forms of inference from propositional logic*.

This leads us to the following result:

If our arithmetic is inconsistent, we can derive a contradiction from correct numerical propositions through the mere application of inferences from propositional logic.

Here "correct numerical propositions" are propositions of the form *One* 1, νPrv , as well as all numerical special cases of general propositions occurring among the axioms such as, e.g., $3 = 3$, $4 = 5 \supset 5 = 4$, $3\text{Pr}4 \supset \neg \text{One } 4$.

It is almost self-evident that from such propositions no contradictions are derivable by means of propositional logic. A proof for this would hardly be more than a formal paraphrasing of an intuitively clear situation of fact. Such a proof will therefore not be carried out save for indicating briefly the customary procedure for it:

We determine generally for which numerical values the formulae *One* μ , $\mu = \nu$, $\mu\text{Pr}\nu$, $\mu + \nu = \rho$, etc., are correct and for which values they are false; furthermore, we explain in the customary way (cf., e.g., Hilbert-Ackermann p. 3) the correctness or falsity of $\mathbf{A} \& \mathbf{B}$, $\mathbf{A} \vee \mathbf{B}$, $\neg \mathbf{A}$, and $\mathbf{A} \supset \mathbf{B}$, as functions of the correctness or falsity of the sub formulae; we then show that all numerical special cases of axiom formulae are "correct"; and finally, that inference figures of propositional logic always lead from correct formulae to other correct formulae. A contradiction formula, however, is not a correct formula.

3.3. It is easy to see from the remarks made in 3.25 in what way the system of *arithmetic axiom formulae* may be extended without making a contradiction derivable in it: We can quite generally allow the introduction of axiom formulae that begin with \forall -symbols spanning the whole formula, which do not contain any \exists -symbols,

and of which every numerical special case is intuitively correct. (We could also admit certain formulae containing \exists -symbols, as long as they can be dealt with in the consistency proof in a way analogous to that of the two cases occurring above.)

E.g., The following axiom formulae for addition are admissible:

$$\begin{aligned} &\forall x \forall y (x\text{Pr}y \supset [x + 1 = y]) \\ &\forall x \forall y \forall z \forall u \forall v ((x\text{Pr}y \& [z + x = u] \& \\ &\quad [z + y = v]) \supset u\text{Pr}v) \\ &\forall x \forall y \forall z \forall u (([x + y = z] \& [x + y = u]) \\ &\quad \supset z = u) \\ &\forall x \forall y \forall z ([x + y = z] \supset [y + x = z]) \\ &\text{etc.} \end{aligned}$$

3.4. Arithmetic without complete induction is, however, of little practical significance, since complete induction is constantly required in number theory. Yet the consistency of arithmetic with complete induction has not been conclusively proved to date.

SECTION V.

THE EQUIVALENCE OF THE NEW CALCULI *Nj*, *NK*, AND *Lj*, *LK* WITH A CALCULUS MODELLED ON THE FORMALISM OF HILBERT

§1.

The Concept of Equivalence

1.1. We shall introduce the following concept of equivalence between *formulae* and *sequents* (which is in harmony with what was said in I, 1.1 and I, 2.4, concerning the intuitive meaning of the symbol Δ and of sequents):

Equal formulae are equivalent.

Equal sequents are equivalent.

Two formulae are equivalent if the replacement of every occurrence of the symbol Δ in one of them by the formula $\Delta \& \neg \Delta$ yields the other formula.

The sequent $\mathbf{A}_1, \dots, \mathbf{A}_\mu \rightarrow \mathbf{B}_1, \dots, \mathbf{B}_\nu$ is equivalent to the following formula:

If the *A*'s and *B*'s are not empty:

$$(\mathbf{A}_1 \& \dots \& \mathbf{A}_\mu) \supset (\mathbf{B}_1 \vee \dots \vee \mathbf{B}_\nu);$$

(this version is more convenient for the equivalence proof than the one with $\mathbf{B}_1 \vee \dots \vee \mathbf{B}_\nu$);

if the *A*'s are empty, but the *B*'s are not:

$B_1 \vee \dots \vee B_n$;

if the B 's are empty, but the A 's are not:

$(A_1 \& \dots \& A_n) \supset (A \& \neg A)$;

if the A 's and the B 's are empty:

$A \& \neg A$.

The equivalence is transitive.

1.2. (We could of course give a considerably wider definition of equivalence, e.g., two formulae are usually called equivalent if one is derivable from the other. Here we shall content ourselves with the particular definition given, which is adequate for our proofs of equivalence.)

Two *derivations* will be called equivalent if the endformula or endsequent of one is equivalent to that of the other.

Two *calculi* will be called equivalent if every derivation in one calculus can be transformed into an equivalent derivation in the other calculus.

In §2 of this section we shall present a calculus (*LDJ* for intuitionist, *LDK* for classical predicate logic) modeled on Hilbert's formalism. In the remaining paragraphs of this section we shall then demonstrate the equivalence of the calculi *LDJ*, *NJ*, and *LJ* (§§3-5) as well as the equivalence of the calculi *LDK*, *NK*, and *LK* (§6) in the sense just explained. We shall thus successively prove the following:

Every *LDJ*-derivation can be transformed into an equivalent *NJ*-derivation (§3); every *NJ*-derivation can be transformed into an equivalent *LJ*-derivation (§4); and every *LJ*-derivation can be transformed into an equivalent *LDJ*-derivation (§5). This obviously proves the equivalence of all three calculi. The three classical calculi are dealt with analogously in §6 (6.1-6.3).

§2.

Presentation of a Logistic Calculus According to Hilbert⁴ and Glivenko⁵

We shall begin by explaining the *intuitionist* form of the calculus:

An *LDJ*-derivation consists of formulae arranged in tree form, where the initial formulae are basic formulae.

The basic formulae and the inference figures are obtained from the following schemata by the same rule of replacement as in II, 2.21, i.e.: For A , B , C , put any arbitrary formula; for $\forall xFx$ or $\exists xFx$ put any arbitrary formula with \forall or \exists for its terminal symbol, where x signifies the relevant bound object variable; for Fa put that formula which results from Fx by the replacement of every occurrence of the bound object variable x by the free object variable a .

Schemata for basic formulae:

- 2.11. $A \supset A$
- 2.12. $A \supset (B \supset A)$
- 2.13. $(A \supset (A \supset B)) \supset (A \supset B)$
- 2.14. $(A \supset (B \supset C)) \supset (B \supset (A \supset C))$
- 2.15. $(A \supset B) \supset ((B \supset C) \supset (A \supset C))$
- 2.21. $(A \& B) \supset A$
- 2.22. $(A \& B) \supset B$
- 2.23. $(A \supset B) \supset ((A \supset C) \supset (A \supset (B \& C)))$
- 2.31. $A \supset (A \vee B)$
- 2.32. $B \supset (A \vee B)$
- 2.33. $(A \supset C) \supset ((B \supset C) \supset ((A \vee B) \supset C))$
- 2.41. $(A \supset B) \supset ((A \supset \neg B) \supset \neg A)$
- 2.42. $(\neg A) \supset (A \supset B)$
- 2.51. $\forall xFx \supset Fa$
- 2.52. $Fa \supset \exists xFx$

(Several of the schemata are dispensable, yet independence does not concern us here.)

Schemata for inference figures:

$\frac{A \quad A \supset B}{B}$	$\frac{A \supset Fa}{A \supset \forall x Fx}$	$\frac{Fa \supset A}{(\exists x Fx) \supset A}$
---------------------------------	---	---

Restriction on Variables: In the inference figures obtained from the last two schemata, the object variable, designated by a in the schema, must not occur in the lower formula (hence not in A and Fx).

(The calculus *LDJ* is essentially equivalent to that of Heyting.⁶)

By including the basic formula schema $A \vee \neg A$, the calculus *LDK* (classical predicate calculus) is obtained.

(This latter calculus is essentially equivalent to the calculus presented in Hilbert-Ackermann, p. 53.)

⁴ D. Hilbert, "Die Grundlagen der Mathematik," *Abhandlungen aus dem mathematischen Seminar der Hamburger Universität*, vol. 6 (1928), pp. 65-85.

⁵ V. Glivenko, "Sur quelques points de la logique de M. Brouwer," *Académie royale de Belgique: Bulletins de la classe des sciences*, 5e. série, tome XV (1929), pp. 183-188.

⁶ Cf. n. 2 of Part I.

§3.

Transformation of an LDJ-Derivation into an Equivalent NJ-Derivation

From an LDJ-derivation (V, 2) we obtain an NJ-derivation (II, 2) with the same endformula by transforming the LDJ-derivation in the following way: (In this transformation all D-formulae of this derivation will reappear as D-formulae of the NJ-derivation, and they will not depend on any assumption formula. Included further will be other D-formulae dependent on assumption formulae.)

3.1. The LDJ-basic formulae are replaced by NJ-derivations according to the following schemata:

$$(2.11) \frac{I \quad A}{A \supset A} \supset -I_1 \quad (2.12) \frac{I \quad A}{B \supset A} \supset -I \quad \frac{B \supset A}{A \supset (B \supset A)} \supset -I_1$$

$$(2.13) \frac{I \quad A \quad A \supset (A \supset B)}{A \supset B} \supset -E \quad \frac{B}{A \supset B} \supset -I_1 \quad \frac{A \supset (A \supset B)}{(A \supset (A \supset B)) \supset (A \supset B)} \supset -I_2$$

$$(2.14) \frac{2 \quad A \quad A \supset (B \supset C)}{B \supset C} \supset -E \quad \frac{C}{A \supset C} \supset -I_1 \quad \frac{B \supset (A \supset C)}{(A \supset (B \supset C)) \supset (B \supset (A \supset C))} \supset -I_3$$

$$(2.15) \frac{I \quad A \quad A \supset B}{B} \supset -E \quad \frac{2 \quad B \supset C}{C} \supset -E \quad \frac{C}{A \supset C} \supset -I_1 \quad \frac{(B \supset C) \supset (A \supset C)}{(A \supset B) \supset ((B \supset C) \supset (A \supset C))} \supset -I_3$$

$$(2.21) \frac{I \quad A \& B}{A} \& -E \quad \frac{(A \& B) \supset A}{(A \& B) \supset A} \supset -I_1$$

2.22, 2.31, 2.32, 2.51 and 2.52 are dealt with analogously to 2.21.

$$(2.23) \frac{I \quad A \quad A \supset B}{B} \supset -E \quad \frac{I \quad A \quad A \supset C}{C} \supset -E \quad \frac{B \& C}{A \supset (B \& C)} \supset -I_1 \quad \frac{(A \supset C) \supset (A \supset (B \& C))}{(A \supset B) \supset ((A \supset C) \supset (A \supset (B \& C)))} \supset -I_3$$

$$(2.33) \frac{I \quad A \quad A \supset C}{A \vee B \quad C} \supset -E \quad \frac{I \quad B \quad B \supset C}{C} \supset -E \quad \frac{C}{(A \vee B) \supset C} \supset -E_2 \quad \frac{(A \vee B) \supset C}{(B \supset C) \supset ((A \vee B) \supset C)} \supset -E_3 \quad \frac{(B \supset C) \supset ((A \vee B) \supset C)}{(A \supset C) \supset ((B \supset C) \supset ((A \vee B) \supset C))} \supset -E_1$$

$$(2.41) \frac{I \quad A \quad A \supset B}{B} \supset -E \quad \frac{I \quad A \quad A \supset \neg B}{\neg B} \supset -E \quad \frac{A}{\neg A} \neg -I_1 \quad \frac{\neg A}{(A \supset \neg B) \supset \neg A} \supset -I_2 \quad \frac{(A \supset \neg B) \supset \neg A}{(A \supset B) \supset ((A \supset \neg B) \supset \neg A)} \supset -I_3$$

$$(2.42) \frac{I \quad A \quad \neg A}{A \quad \neg A} \neg -E \quad \frac{A}{\neg A} \neg -I_1 \quad \frac{A \supset B}{(\neg A) \supset (A \supset B)} \supset -I_3$$

3.2. The LDJ-inference figures are replaced by sections of an NJ-derivation according to the following schemata:

$\frac{A \quad A \supset B}{B}$ remains as it is, since it has already the form of a $\supset -E$.

$$\frac{A \supset Fa}{A \supset \forall x Fx} \text{ becomes: } \frac{I \quad A \quad A \supset Fa}{Fa} \supset -E \quad \frac{Fa}{\forall x Fx} \forall -I \quad \frac{\forall x Fx}{A \supset \forall x Fx} \supset -I_1$$

$$\frac{Fa \supset A}{(\exists x Fx) \supset A} \text{ becomes: } \frac{2 \quad Fa \quad Fa \supset A}{\exists x Fx \quad A} \supset -E \quad \frac{A}{(\exists x Fx) \supset A} \supset -I_1$$

The restriction on variables for \forall -I and \exists -E is fulfilled, as is easily seen, by virtue of the restriction on variables existing for $LD\mathcal{J}$ -inference figures.

This completes the transformation of an $LD\mathcal{J}$ -derivation into an equivalent $N\mathcal{J}$ -derivation.

§4.

Transformation of an $N\mathcal{J}$ -derivation into an Equivalent $L\mathcal{J}$ -derivation

4.1. We proceed as follows: First we *replace* every D -formula of the $N\mathcal{J}$ -derivation by the following sequent (cf. III, 1.1): In its succedent only the formula itself occurs; in its antecedent occur the assumption formulae upon which the sequent depended, and they occur in the same order from left to right as they did in the $N\mathcal{J}$ -derivation. (It is presumably clear what is meant by the order from left to right of the initial formulae of a figure in tree form.)

We then replace every occurrence of the symbol A by $A \& \neg A$. (The formula resulting from A in this way will be designated by A^* .)

4.2. Now we have already a system of sequents in tree form. The antecedent of the endsequent is empty (II, 2.2); it is obviously already equivalent to the endsequent of the $N\mathcal{J}$ -derivation. The initial sequents all have the form $D^* \rightarrow D^*$ (II, 2.2) and are thus already basic sequents of an $L\mathcal{J}$ -derivation.

The figures formed from $N\mathcal{J}$ -inference figures are transformed into sections of an $L\mathcal{J}$ -derivation according to the following schemata:

4.21. The inference figures \vee -I, \forall -I, and \exists -I have already become $L\mathcal{J}$ -inference figures as a result of the substitution performed. (In the case of a \forall -I, the $L\mathcal{J}$ -restriction on variables is fulfilled by virtue of the $N\mathcal{J}$ -restriction on variables.)

4.22. $A \& \neg I$ became:

$$\frac{\Gamma \rightarrow A^* \quad \Delta \rightarrow B^*}{\Gamma, \Delta \rightarrow A^* \& B^*}$$

This is transformed into:

$$\frac{\frac{\Gamma \rightarrow A^*}{\Gamma, \Delta \rightarrow A^*} \text{ possibly several interchanges and contractions} \quad \frac{\Delta \rightarrow B^*}{\Gamma, \Delta \rightarrow B^*} \text{ possibly several thinnings \&-IS}}{\Gamma, \Delta \rightarrow A^* \& B^*}$$

4.23. $A \supset$ -I became

$$\frac{\Gamma_1, A^*, \Gamma_2, \dots, A^*, \Gamma_p \rightarrow B^*}{\Gamma_1, \Gamma_2, \dots, \Gamma_p \rightarrow A^* \supset B^*}$$

This we transform into:

$$\frac{\Gamma_1, A^*, \Gamma_2, \dots, A^*, \Gamma_p \rightarrow B^*}{\Gamma_1, \Gamma_2, \dots, \Gamma_p \rightarrow A^* \supset B^*} \begin{array}{l} \text{possibly several interchanges and contractions, sometimes a thinning} \\ \supset\text{-IS} \end{array}$$

4.24. The same procedure applies to a \neg -I. Finally, we still have to consider the figure

$$\frac{A^*, \Gamma \rightarrow A \& \neg A}{\Gamma \rightarrow \neg A^*}$$

First we derive $A \& \neg A \rightarrow$ in the calculus $L\mathcal{J}$ as follows:

$$\frac{\frac{\frac{A \rightarrow A}{\neg A, A \rightarrow} \neg\text{-IA}}{A \& \neg A, A \rightarrow} \&\text{-IA}}{\frac{A, A \& \neg A \rightarrow}{A \& \neg A, A \& \neg A \rightarrow} \text{interchange}} \&\text{-IA} \text{ contraction}$$

By including this sequent, the figure in question is transformed as follows:

$$\frac{\frac{A^*, \Gamma \rightarrow A \& \neg A}{A^*, \Gamma \rightarrow} \quad \frac{A \& \neg A \rightarrow}{\Gamma \rightarrow \neg A^*} \text{ cut}}{\Gamma \rightarrow \neg A^*} \neg\text{-IS}$$

4.25. By substitution (4.1) the $N\mathcal{J}$ -inference figure $\frac{A}{D}$ became:

$$\frac{\Gamma \rightarrow A \& \neg A}{\Gamma \rightarrow D^*}$$

This is transformed into:

$$\frac{\frac{\Gamma \rightarrow A \& \neg A}{\Gamma \rightarrow} \quad \frac{A \& \neg A \rightarrow}{\Gamma \rightarrow D^*} \text{ cut}}{\Gamma \rightarrow D^*} \text{ thinning}$$

The derivation for $A \& \neg A \rightarrow$, as presented in 4.24, should here still be written above that sequent.

$$4.26. A \forall\text{-E became: } \frac{\Gamma \rightarrow \forall x F^* x}{\Gamma \rightarrow F^* a}$$

This is transformed into:

$$\frac{\frac{\Gamma \rightarrow \forall x F^* x}{\Gamma \rightarrow F^* a} \quad \frac{F^* a \rightarrow F^* a}{\forall x F^* x \rightarrow F^* a} \forall\text{-IA}}{\Gamma \rightarrow F^* a} \text{ cut}$$

4.27. The same method is used for $\&-E$.

4.28. $A \supset -E$ became:

$$\frac{\Gamma \rightarrow A^* \quad \Delta \rightarrow A^* \supset B^*}{\Gamma, \Delta \rightarrow B^*}$$

This is transformed into:

$$\frac{\Delta \rightarrow A^* \supset B^* \quad \frac{\Gamma \rightarrow A^* \quad B^* \rightarrow B^*}{A^* \supset B^*, \Gamma \rightarrow B^*} \supset -IA}{\frac{\Delta, \Gamma \rightarrow B^*}{\Gamma, \Delta \rightarrow B^*} \text{ possibly several interchanges}} \text{ cut}$$

4.29. $A \neg -E$ became:

$$\frac{\Gamma \rightarrow A^* \quad \Delta \rightarrow \neg A^*}{\Gamma, \Delta \rightarrow A \& \neg A}$$

This is transformed into:

$$\frac{\Delta \rightarrow \neg A^* \quad \frac{\Gamma \rightarrow A^*}{\neg A^*, \Gamma \rightarrow} \neg -IA}{\frac{\Delta, \Gamma \rightarrow}{\Gamma, \Delta \rightarrow} \text{ possibly several interchanges}} \text{ cut}$$

thinning

4.210. $\vee -E$. Both righthand upper sequents are to be followed, as in the case of a $\supset -I$ and $\neg -I$ (4.23) above, by interchanges, contractions, and thinnings (wherever necessary) so that in each case the result is a sequent in whose antecedent occurs a formula of the form A^* or B^* at the beginning (whereas the original assumption formulae involved have disappeared precisely into the rest of the antecedent). Then follows:

$$\frac{\frac{A^*, \Gamma \rightarrow C^* \text{ possibly several thinnings and } B^*, \Delta \rightarrow C^* \text{ possibly several thinnings and}}{A^*, \Gamma, \Delta \rightarrow C^* \text{ interchanges } B^*, \Gamma, \Delta \rightarrow C^* \text{ interchanges}} \vee -IA}{\frac{E \rightarrow A^* \vee B^* \quad A^* \vee B^*, \Gamma, \Delta \rightarrow C^*}{E, \Gamma, \Delta \rightarrow C^*} \text{ cut}}$$

4.211. $A \exists -E$ is treated quite similarly: First we move F^*a in the righthand upper sequent to the beginning of the antecedent (cf. 4.23); then follows:

$$\frac{\Delta \rightarrow \exists x F^*x \quad \frac{F^*a, \Gamma \rightarrow C^*}{\exists x F^*x, \Gamma \rightarrow C^*} \exists -IA}{\Delta, \Gamma \rightarrow C^*} \text{ cut}$$

The LJ -restriction on variables for $\exists -IA$ is fulfilled by virtue of the NJ -restriction on variables for $\exists -E$.

This completes the transformation of an NJ -derivation into an equivalent LJ -derivation.

§5.

Transformation of an LJ -derivation into an Equivalent LDJ -derivation

This transformation is a little more difficult than the two previous ones. We shall carry it out in a number of distinct steps.

Preliminary Remark: Contractions and interchanges in the succedent do not occur in the calculus LJ , since they require the occurrence of at least two S -formulae in the succedent.

5.1. We first introduce *new basic sequents* in place of the figures $\&-IA$, $\vee -IS$, $\neg -IA$, $\exists -IS$, $\neg -IA$, and $\supset -IA$; these are to be formed according to the following schemata (rule of replacement as in III, 1.2—the same rule will always apply below; in addition to the letters A , B , D , and E we shall also, incidentally, use the letters C , H , and J):

Bs1: $A \& B \rightarrow A$	Bs2: $A \& B \rightarrow B$
Bs3: $A \rightarrow A \vee B$	Bs4: $B \rightarrow A \vee B$
Bs5: $\neg x F x \rightarrow F a$	Bs6: $F a \rightarrow \exists x F x$
Bs7: $\neg A, A \rightarrow$	Bs8: $A \supset B, A \rightarrow B$

Thus in the LJ -derivation to be considered, we transform the inference figures concerned in the following way:

$A \&-IA$ becomes:

$$\frac{\text{Bs1} \quad A \& B \rightarrow A \quad A, \Gamma \rightarrow \Theta}{A \& B, \Gamma \rightarrow \Theta} \text{ cut}$$

The other form of the $\&-IA$ is transformed correspondingly, so is every $\&-IA$.

$\vee -IS$ and $\exists -IS$ are dealt with symmetrically.

$A \neg -IA$ becomes:

$$\frac{\Gamma \rightarrow A \quad \frac{\text{Bs7} \quad \neg A, A \rightarrow}{A, \neg A \rightarrow} \text{ interchange}}{\frac{\Gamma, \neg A \rightarrow}{\neg A, \Gamma \rightarrow} \text{ possibly several interchanges}} \text{ cut}$$

(The Θ in the schema of $\neg -IA$ (III, 1.22) must

be empty by virtue of the $L\bar{J}$ -condition for succedents; the same holds for the $\supset -IA$.)

$A \supset -IA$ becomes:

$$\frac{\Gamma \rightarrow A \quad \frac{\text{Bs8} \quad \frac{A \supset B, A \rightarrow B}{A, A \supset B \rightarrow B} \text{interchange}}{\Gamma, A \supset B \rightarrow B} \text{cut}}{\Gamma, A \supset B, \Delta \rightarrow A} \text{cut} \quad \frac{B, \Delta \rightarrow A}{\Gamma, A \supset B, \Delta \rightarrow A} \text{possibly several interchanges}$$

5.2. We now write the formula $A \& \neg A$ in the succedent of all D -formulae whose *succedent is empty*.

In doing so the basic sequents of the form $D \rightarrow D$, as well as **Bs1** to **Bs6** and **Bs8**, also the figures $\&-IS$, $\vee-IS$, and $\supset-IS$, remain unchanged. The other basic sequents and inference figures are transformed into new basic sequents and inference figures according to the following schemata:

Bs9: $A, \neg A \rightarrow B$

$$\text{If1: } \frac{\Gamma \rightarrow H}{D, \Gamma \rightarrow H} \quad \text{If2: } \frac{D, D, \Gamma \rightarrow H}{D, \Gamma \rightarrow H} \quad \text{If3: } \frac{\Delta, D, E, \Gamma \rightarrow H}{\Delta, E, D, \Gamma \rightarrow H}$$

$$\text{If4: } \frac{\Gamma \rightarrow A \& \neg A}{\Gamma \rightarrow D} \quad \text{If5: } \frac{\Gamma \rightarrow D \quad D, \Delta \rightarrow H}{\Gamma, \Delta \rightarrow H}$$

$$\text{If6: } \frac{A, \Gamma \rightarrow H \quad B, \Gamma \rightarrow H}{A \vee B, \Gamma \rightarrow H} \quad \text{If7: } \frac{Fa, \Gamma \rightarrow H}{\exists x Fx, \Gamma \rightarrow H}$$

$$\text{If8: } \frac{A, \Gamma \rightarrow A \& \neg A}{\Gamma \rightarrow \neg A}$$

(for **If7** there exists the following restriction on variables: The free object variable designated by a must not occur in the lower sequent.)

5.3. Now the inference figure **If4** may be replaced by other figures as follows (this is mainly due to our having kept general the form of the schema **Bs9**):

$$\frac{\text{Bs2} \quad \frac{\Gamma \rightarrow A \& \neg A \quad A \& \neg A \rightarrow \neg A}{\Gamma \rightarrow \neg A} \text{Bs1} \quad \frac{\Gamma \rightarrow \neg A}{\Gamma \rightarrow A} \text{Bs9} \quad \frac{\neg A, A \rightarrow D}{\Gamma, A \rightarrow D} \text{If5}}{\Gamma, \Gamma \rightarrow D} \text{possibly several If3's} \quad \frac{\Gamma, \Gamma \rightarrow D}{\Gamma \rightarrow D} \text{possibly several If2's and If3's}$$

In a similar way we replace the inference figure **If8** (wherever it occurs in the derivation), only this

time we use a new inference figure according to the following schema:

$$\text{If9: } \frac{\Gamma, A \rightarrow A \quad \Gamma, A \rightarrow \neg A}{\Gamma \rightarrow \neg A}$$

We substitute as follows (in place of **If8** we obtain):

$$\frac{\text{Bs1} \quad \frac{A, \Gamma \rightarrow A \& \neg A \quad A \& \neg A \rightarrow A}{A, \Gamma \rightarrow A} \text{If3} \quad \frac{A, \Gamma \rightarrow A}{\Gamma, A \rightarrow A} \text{possibly several If3's} \quad \text{Bs2} \quad \frac{A, \Gamma \rightarrow A \& \neg A \quad A \& \neg A \rightarrow \neg A}{A, \Gamma \rightarrow \neg A} \text{If5} \quad \frac{A, \Gamma \rightarrow \neg A}{\Gamma, A \rightarrow \neg A} \text{possibly several If3's}}{\Gamma \rightarrow \neg A} \text{If9}$$

5.4. We now introduce two further inference figure schemata, viz.:

$$\text{If10: } \frac{\Gamma, A \rightarrow B}{\Gamma \rightarrow A \supset B} \quad \text{and its converse: } \text{If11: } \frac{\Gamma \rightarrow A \supset B}{\Gamma, A \rightarrow B}$$

These two types of inference figures are introduced into the derivation in order to enable us to replace a number of other inference figures by more specialized ones (in 5.42 and 5.43).

5.41. To begin with, $\supset-IS$ inference figures are now replaceable by means of **If10**:

$A \supset -IS$ is transformed into:

$$\frac{A, \Gamma \rightarrow B}{\Gamma, A \rightarrow B} \text{possibly several If3's} \quad \frac{\Gamma, A \rightarrow B}{\Gamma \rightarrow A \supset B} \text{If10}$$

5.42. The inference figures **If1**, **If2**, **If3**, **If5**, **If6**, and **If7** are then transformed in the following way:

As an example we take an **If2**, which is transformed into the following figure (suppose Γ equals J_1, \dots, J_p):

$$\frac{\frac{D, D, J_1, \dots, J_p \rightarrow H}{D, D, J_1, \dots, J_{p-1} \rightarrow J_p \supset H} \text{If10} \quad \frac{D, D \rightarrow J_1 \supset (J_2 \supset \dots (J_p \supset H).)}{D \rightarrow J_1 \supset (J_2 \supset \dots (J_p \supset H).)} \text{several If10's}}{D, J_1, \dots, J_p \rightarrow H} \text{several If11's}$$

We proceed quite analogously with all other figures mentioned, i.e., using **If10** and **If11**, we replace them by inference figures according to these schemata:

$$\text{If12: } \frac{\rightarrow C}{D \rightarrow C} \quad \text{If13: } \frac{D, D \rightarrow C}{D \rightarrow C} \quad \text{If14: } \frac{\Delta, D, E \rightarrow C}{\Delta, E, D \rightarrow C} \quad \text{If15: } \frac{\Gamma \rightarrow D \quad D \rightarrow C}{\Gamma \rightarrow C} \quad \text{If16: } \frac{A \rightarrow C \quad B \rightarrow C}{A \vee B \rightarrow C} \quad \text{If17: } \frac{Fa \rightarrow C}{\exists x Fx \rightarrow C}$$

(for **If**₁₇ there exists a restriction on variables: The free object variable designated by **a** must not occur in the lower sequent.)

5.43. In a similar way we also replace the inference figures **If**₉, **If**₁₃, and **If**₁₄ by the following (using **If**₁₀ and **If**₁₁):

$$\text{If}_{18}: \frac{\Gamma \rightarrow A \supset A \quad \Gamma \rightarrow A \supset \neg A}{\Gamma \rightarrow \neg A} \quad \text{If}_{19}: \frac{\rightarrow D \supset (D \supset C)}{\rightarrow D \supset C}$$

$$\text{If}_{20}: \frac{A \rightarrow D \supset (E \supset C)}{A \rightarrow E \supset (D \supset C)}$$

The basic sequents **Bs**₈ and **Bs**₉ may be replaced in the same way by:

$A \supset B \rightarrow A \supset B$, this form falls under the schema $D \rightarrow D$; as well as **Bs**₁₀: $\neg A \rightarrow A \supset H$.

5.5. Now comes the *final step*:

Every D -sequent

$$A_1, \dots, A_n \rightarrow B$$

is replaced by the formula $(A_1 \& \dots \& A_n) \supset B$.

(If the A 's are empty, we mean **B**. An empty succedent no longer occurs, according to 5.2.)

All basic sequents (viz. $D \rightarrow D$, **Bs**₁ to **Bs**₆, **Bs**₁₀) are thus transformed into $LD\mathcal{J}$ -basic sequents.

Of the inference figures, \forall - IS and **If**₁₇ are also transformed into $LD\mathcal{J}$ -inference figures. (\forall - IS , however, forms an exception if Γ is empty. In that case we first derive (in the $LD\mathcal{J}$ -calculus) $(A \supset A) \supset Fa$ from **Fa** by means of 2.12, and by then applying the $LD\mathcal{J}$ -inference figure, we finally obtain $\forall xFx$ once more by means of 2.11.)

The figures obtained from the remaining inference figures (which are $\&$ - IS , **If**₁₀, 11, 12, 15, 16, 18, 19, 20) by substitution, are turned into sections of an $LD\mathcal{J}$ -derivation in the following way:

A $\&$ - IS has become (suppose first that Γ is not empty):

$$\begin{array}{c} \frac{C \supset A \quad C \supset B}{C \supset (A \& B)} \quad \text{This is transformed into:} \\ \frac{C \supset A \quad (C \supset A) \supset ((C \supset B) \supset (C \supset (A \& B)))}{C \supset B \quad (C \supset B) \supset (C \supset (A \& B))} \\ \hline C \supset (A \& B) \end{array}$$

If Γ is empty, we proceed as in the case of \forall - IS .

The figures obtained from **If**₁₂, 15, 16, and 19 by substitution may be dealt with quite analogously using basic formulae according to the schemata 2.12, 2.15, 2.33, and 2.13.

In a similar way **If**₁₈ and **If**₂₀ are dealt with by means of 2.41 and 2.14 and the application of 2.15 and 2.14, 2.13.

The only figures now left are those having resulted from **If**₁₀ and **If**₁₁. Both are trivial for an empty Γ , hence, below, suppose that Γ is not empty. In that case we transform these figures into sections of $LD\mathcal{J}$ -derivations as follows:

(**If**₁₀): From $(C \& A) \supset B$ we have to derive $C \supset (A \supset B)$. Now 2.23 together with 2.11 yields: $(C \supset A) \supset (C \supset (C \& A))$. This together with $(C \& A) \supset B$ and 2.15, 2.14 yields $(C \supset A) \supset (C \supset B)$, and from this formula together with 2.12, 2.15, we obtain $A \supset (C \supset B)$, and by 2.14 $C \supset (A \supset B)$ results.

(**If**₁₁): From $C \supset (A \supset B)$ we derive $(C \& A) \supset B$ in the $LD\mathcal{J}$ -calculus as follows: 2.21 and 2.22 yield $(C \& A) \supset C$ and $(C \& A) \supset A$; and from this together with $C \supset (A \supset B)$ we obtain $(C \& A) \supset B$ (by using 2.15, 2.14, 2.15, 2.13).

This completes the transformation of the $L\mathcal{J}$ -derivation into an $LD\mathcal{J}$ -derivation. Furthermore, the two derivations really are equivalent, since the endsequent of the $L\mathcal{J}$ -derivation was affected only by the transformations 5.2 and 5.5, and has thus obviously been transformed into a formula equivalent with it (according to 1.1).

If the results of §§3–5 are taken together, the *equivalence of the three calculi* $LD\mathcal{J}$, $N\mathcal{J}$, and $L\mathcal{J}$ is now fully proved.

§6.

The Equivalence of the Calculi LDK, NK, and LK

Once the equivalence of the different *intuitionist* calculi has been proved, it is fairly easy to deduce that of the *classical* calculi.

6.1. In order to transform an LDK -derivation into an equivalent NK -derivation we proceed exactly as in §3. The additional basic formulae according to the schema $A \vee \neg A$ remain unchanged, and are thus at once basic formulae of the NK -derivation.

6.2. In order to transform an NK -derivation into an equivalent LK -derivation we proceed initially as in §4. In doing so the additional basic formulae according to the schema $A \vee \neg A$ are transformed into sequents of the form $\rightarrow A^* \& \neg A^*$. These we then replace by their LK -derivations (according to III, 1.4). The transformation of an

NK-derivation into an equivalent *LK*-derivation is thus completed.

6.3. *Transformation of an LK-derivation into an LDK-derivation.*

We introduce an *auxiliary calculus* differing from the *LK*-calculus in the following respect:

Inference figures may be formed according to the schemata III, 1.21, 1.22, yet with the following restrictions: Contractions and interchanges in the succedent are not permissible; in the remaining schemata no substitution may be performed on Θ and Δ ; these places thus remain empty.

Furthermore, the following two schemata for inference figures are added (rule of replacement as usual: III, 1.2):

$$\text{If1: } \frac{\Gamma \rightarrow \mathbf{A}, \Theta}{\Gamma, \neg \mathbf{A} \rightarrow \Theta} \text{ and its converse: } \text{If2: } \frac{\Gamma, \neg \mathbf{A} \rightarrow \Theta}{\Gamma \rightarrow \mathbf{A}, \Theta}$$

(Here Θ need thus not be empty.)

6.31. *Transformation of an LK-derivation into a derivation of the auxiliary calculus:*

(The procedure is similar to that in 5.4.)

All inference figures, with the exception of contractions and interchanges in the succedent, are transformed according to the following rule: The upper sequents are followed by inference figures **If1**, until all formulae of Θ or Δ have been negated and brought into the antecedent (to the right of Γ or Δ). Then follows an inference figure of the same kind as the one just transformed, which is now actually a permissible inference figure in the auxiliary calculus. (The formulae that have been brought into the antecedent are treated as part of Γ or Δ .) Then follow **If2** inference figures, and Θ and Δ are thus brought back into the succedent. (In the case of a $\supset -IA$ and a cut, we may possibly have to carry out interchanges in the antecedent, but these are also permissible inference figures in the auxiliary calculus.)

Now we still have to consider contractions—or interchanges—in the succedent. Here, as in the previous case, the *whole* succedent is negated and brought forward into the antecedent. We then carry out interchanges, a contraction, and further interchanges—or one interchange—in the antecedent, and then the negated formulae are brought back into the succedent (by means of the inference figures **If2**).

6.32. Transformation of a derivation of the auxiliary calculus into a derivation of the calculus *LJ* augmented by the inclusion of the basic sequent schema $\rightarrow \mathbf{A} \vee \neg \mathbf{A}$:

We begin by transforming all *D*-sequents as follows:

$\mathbf{A}_1, \dots, \mathbf{A}_\mu \rightarrow \mathbf{B}_1, \dots, \mathbf{B}_\nu$ becomes:

$\mathbf{A}_1, \dots, \mathbf{A}_\mu \rightarrow \mathbf{B}_\nu \vee \dots \vee \mathbf{B}_1$. If the succedent is empty, it remains empty.

Now all basic sequents or inference figures of the auxiliary calculus with the exception of the figures **If1** and **If2**, have thus already become basic sequents or inference figures of the calculus *LJ*. This is so since these inference figures have resulted from the schemata III, 1.21, 1.22 (with the exception of the schemata for contraction and interchange in the succedent) in that Θ and Δ have always remained empty. Hence at most one formula could have occurred in the succedent.

Hence we still have to transform the figures which have resulted from the inference figures **If1** and **If2** in the course of the above modification.

6.321. First **If1**: If Θ is empty, we replace the inference figure by an $\neg -IA$ to be followed by interchanges in the antecedent. Suppose Θ is not empty, where Θ^* designates the formulae belonging to Θ , in reverse order and connected by \vee .

In that case the inference figure runs as follows after the transformation of the succedents:

$$\frac{\Gamma \rightarrow \Theta^* \vee \mathbf{A}}{\Gamma, \neg \mathbf{A} \rightarrow \Theta^*}$$

This is transformed into the following section of an *LJ*-derivation:

$$\frac{\frac{\frac{\frac{\Theta^* \rightarrow \Theta^*}{\neg \mathbf{A}, \Theta^* \rightarrow \Theta^*} \text{ thinning}}{\Theta^*, \neg \mathbf{A} \rightarrow \Theta^*} \text{ interchange}}{\Gamma \rightarrow \Theta^* \vee \mathbf{A}} \quad \frac{\frac{\frac{\mathbf{A} \rightarrow \mathbf{A}}{\neg \mathbf{A}, \mathbf{A} \rightarrow} \neg -IA}{\mathbf{A}, \neg \mathbf{A} \rightarrow} \text{ interchange}}{\mathbf{A}, \neg \mathbf{A} \rightarrow \Theta^*} \text{ thinning}}{\Gamma, \neg \mathbf{A} \rightarrow \Theta^*} \text{ cut}$$

6.322. An inference figure **If2** runs as follows after the transformation of its succedents:

$$\frac{\Gamma, \neg \mathbf{A} \rightarrow \Theta^*}{\Gamma \rightarrow \Theta^* \vee \mathbf{A}}$$

where Θ^* has the same meaning as in the previous case. If Θ is empty, assume Θ^* to be empty too, and let $\Theta^* \vee \mathbf{A}$ mean \mathbf{A} .

It is transformed into the following section of a derivation:

V. FAMILY RESEMBLANCES AND GENERALIZATION CONCERNING THE ARTS

MAURICE MANDELBAUM

IN 1954 William Elton collected and published a group of essays under the title *Aesthetics and Language*. As his introduction made clear, a common feature of these essays was the application to aesthetic problems of some of the doctrines characteristic of recent British linguistic philosophy.¹ While this mode of philosophizing has not had as pervasive an influence on aesthetics as it has had on most other branches of philosophy,² there have been a number of important articles which, in addition to those contained in the Elton volume, suggest the direction in which this influence runs. Among these articles one might mention "The Task of Defining a Work of Art" by Paul Ziff,³ "The Role of Theory in Aesthetics" by Morris Weitz,⁴ Charles L. Stevenson's "On 'What is a Poem'"⁵ and W. E. Kennick's "Does Traditional Aesthetics Rest on a Mistake?"⁶ In each of them one finds a conviction which was also present in most of the essays in the Elton volume: that it is a mistake to offer generalizations concerning the arts, or, to put the matter in a more provocative manner, that it is a mistake to attempt to discuss what art, or beauty, or the aesthetic, or a poem, *essentially* is. In partial support of this contention, some writers have made explicit use of Wittgenstein's doctrine of *family resemblances*; Morris Weitz, for example, has placed it in the forefront of his discussion. However, in that influential and frequently anthologized article, Professor Weitz made no attempt to analyze, clarify, or defend the doctrine itself. Since its use with respect to aesthetics has provided the means by which others have

sought to escape the need of generalizing concerning the arts, I shall begin my discussion with a consideration of it.

I

The *locus classicus* for Wittgenstein's doctrine of family resemblances is in Part I of *Philosophical Investigations*, sections 65-77.⁷ In discussing what he refers to as language-games, Wittgenstein says:

Instead of producing something common to all that we call language, I am saying that these phenomena have no one thing in common which makes us use the same word for all—but they are *related* to one another in many different ways. And it is because of this relationship, or these relationships, that we call them all "language." (§65)

He then illustrates his contention by citing a variety of *games*, such as board games, card games, ball games, etc., and concludes:

We see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities of detail. (§66)

I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc., etc. overlap and criss-cross in the same way.—And I shall say: "games" form a family. (§67)

In short, what Wittgenstein aims to establish is that one need not suppose that all instances of those

¹ See William Elton (ed.), *Aesthetics and Language* (Oxford, Basil Blackwell, 1954), p. 1, n. 1 and 2.

² A discussion of this fact is to be found in Jerome Stolnitz, "Notes on Analytic Philosophy and Aesthetics," *British Journal of Aesthetics*, vol. 3 (1961), pp. 210-222.

³ *Philosophical Review*, vol. 62 (1953), pp. 58-78.

⁴ *Journal of Aesthetics and Art Criticism*, vol. 15 (1956), pp. 27-35.

⁵ *Philosophical Review*, vol. 66 (1957), pp. 329-362.

⁶ *Mind*, vol. 67 (1958), pp. 317-334. In addition to the articles already referred to, I might mention "The Uses of Works of Art" by Teddy Brunius in *Journal of Aesthetics and Art Criticism*, vol. 22 (1963), pp. 123-133, which refers to both Weitz and Kennick, but raises other question with which I am not here concerned.

⁷ Ludwig Wittgenstein, *Philosophical Investigations*, translated by G. E. M. Anscombe (New York, Macmillan, 1953), pp. 31-36. A parallel passage is to be found in "The Blue Book": see *Preliminary Studies for the "Philosophical Investigations," Generally Known as The Blue and Brown Books* (Oxford, Basil Blackwell, 1958), pp. 17-18.

entities to which we apply a common name do in fact possess any one feature in common. Instead, the use of a common name is grounded in the criss-crossing and overlapping of resembling features among otherwise heterogeneous objects and activities.

Wittgenstein's concrete illustrations of the diversity among various types of games may at first make his doctrine of family resemblances extremely plausible. For example, we do not hesitate to characterize tennis, chess, bridge, and solitaire as games, even though a comparison of them fails to reveal any specific feature which is the same in each of them. Nonetheless, I do not believe that his doctrine of family resemblances, as it stands, provides an adequate analysis of why a common name, such as "a game," is in all cases applied or withheld.

Consider first the following case. Let us assume that you know how to play that form of solitaire called "Canfield"; suppose also that you are acquainted with a number of other varieties of solitaire (Wittgenstein uses "patience," i.e., "solitaire," as one instance of a form of game). Were you to see me shuffling a pack of cards, arranging the cards in piles, some face up and some face down, turning cards over one-by-one, sometimes placing them in one pile, then another, shifting piles, etc., you might say: "I see you are playing cards. What game are you playing?" However, to this I might answer: "I am not playing a game; I am telling (or reading) fortunes." Will the resemblances between what you have seen me doing and the characteristics of card games with which you are familiar permit you to contradict me and say that I am indeed playing some sort of game? Ordinary usage would not, I believe, sanction our

describing fortune-telling as an example of playing a game, no matter how striking may be the resemblances between the ways in which cards are handled in playing solitaire and in telling fortunes. Or, to choose another example, we may say that while certain forms of wrestling contests are sometimes characterized as games (Wittgenstein mentions "*Kampfspiele*")⁸ an angry struggle between two boys, each trying to make the other give in, is not to be characterized as a game. Yet one can find a great many resembling features between such a struggle and a wrestling match in a gymnasium. What would seem to be crucial in our designation of an activity as a game is, therefore, not merely a matter of noting a number of specific resemblances between it and other activities which we denote as games, but involves something further.

To suggest what sort of characteristic this "something further" might possibly be, it will be helpful to pay closer attention to the notion of what constitutes a family resemblance. Suppose that you are shown ten or a dozen photographs and you are then asked to decide which among them exhibit strong resemblances.⁹ You might have no difficulty in selecting, say, three of the photographs in which the subjects were markedly round-headed, had a strongly prognathous profile, rather deep-set eyes, and dark curly hair.¹⁰ In some extended, metaphorical sense you might say that the similarities in their features constituted a family resemblance among them. The sense, however, would be metaphorical, since in the absence of a biological kinship of a certain degree of proximity we would be inclined to speak only of resemblances, and not of a *family* resemblance. What marks the difference between a literal and a metaphorical sense of the

⁸ Ludwig Wittgenstein, *Philosophical Investigations*, §66, p. 31. For reasons which are obscure, Miss Anscombe translates "*Kampfspiele*" as "Olympic games."

⁹ In an article which is closely related to my discussion, but which uses different arguments to support a similar point, Haig Khatchadourian has shown that Wittgenstein is less explicit than he should have been with respect to the levels of determinateness at which these resemblances are significant for our use of common names. See "Common Names and 'Family Resemblances'," *Philosophy and Phenomenological Research*, vol. 18 (1957-58), pp. 341-358. (For a related, but less closely relevant article by Professor Khatchadourian see "Art-Names and Aesthetic Judgments," *Philosophy*, vol. 36 [1961], pp. 30-48.)

¹⁰ It is to be noted that this constitutes a closer resemblance than that involved in what Wittgenstein calls "family resemblances," since in my illustration the specific similarities all pertain to a single set of features, with respect to each one of which all three of the subjects directly resemble one another. In Wittgenstein's use of the notion of family resemblances there is, however, no one set of resembling features common to each member of the "family"; there is merely a criss-crossing and overlapping among the elements which constitute the resemblances among the various persons. Thus, in order to conform to his usage, my illustration would have to be made more complicated, and the degree of resemblance would become more attenuated. For example, we would have to introduce the photographs of other subjects in which, for example, recessive chins would supplant prognathous profiles among those who shared the other characteristics; some would have blond instead of dark hair, and protruberant instead of deep-set eyes, but would in each case resemble the others in other respects, etc. However, if what I say concerning family resemblances holds of the stronger similarities present in my illustration, it should hold *a fortiori* of the weaker form of family resemblances to which Wittgenstein draws our attention.

notion of "family resemblances" is, therefore, the existence of a genetic connection in the former case and not in the latter. Wittgenstein, however, failed to make explicit the fact that the literal, root notion of a family resemblance includes this genetic connection no less than it includes the existence of noticeable physiognomic resemblances.¹¹ Had the existence of such a *twofold* criterion been made explicit by him, he would have noted that there is in fact an attribute common to all who bear a family resemblance to each other: they are related through a common ancestry. Such a relationship is not, of course, one among the specific features of those who share a family resemblance; it nonetheless differentiates them from those who are not to be regarded as members of a single family.¹² If, then, it is possible that the analogy of family resemblances could tell us something about how games may be related to one another, one should explore the possibility that, in spite of their great dissimilarities, games may possess a common attribute which, like biological connection, is not itself one among their directly exhibited characteristics. Unfortunately, such a possibility was not explored by Wittgenstein.

To be sure, Wittgenstein does not explicitly state that the resemblances which are correlated with our use of common names must be of a sort that are

directly exhibited. Nonetheless, all of his illustrations in the relevant passages involve aspects of games which would be included in a description of how a particular game is to be played; that is, when he commands us to "look and see" whether there is anything common to all games,¹³ the "anything" is taken to represent precisely the sort of manifest feature that is described in rule-books, such as Hoyle. However, as we have seen in the case of family resemblances, what constitutes a *family* is not defined in terms of the manifest features of a random group of people; we must first characterize the *family* relationship in terms of genetic ties, and then observe to what extent those who are connected in this way *resemble* one another.¹⁴ In the case of games, the analogue to genetic ties might be the purpose for the sake of which various games were formulated by those who invented or modified them, e.g., the potentiality of a game to be of absorbing non-practical interest to either participants or spectators. If there were any such common feature one would not expect it to be defined in a rule book, such as Hoyle, since rule books only attempt to tell us how to play a particular game: our interest in playing a game, and our understanding of what constitutes a game, is already presupposed by the authors of such books.

¹¹ Although Wittgenstein failed to make explicit the fact that a genetic connection was involved in his notion of "family resemblances," I think that he did in fact presuppose such a connection. If I am not mistaken, the original German makes this clearer than does the Anscombe translation. The German text reads:

Ich kann diese Ähnlichkeiten nicht besser charakterisieren, als durch das Wort "Familienähnlichkeiten"; denn so übergreifen und kreuzen sich die verschiedenen Ähnlichkeiten, die zwischen den Gliedern einer Familie bestehen: Wuchs, Gesichtszüge, Augenfarbe, Gang, Temperament, etc., etc. (§67).

Modifying Miss Anscombe's translation in as few respects as possible, I suggest that a translation of this passage might read:

I can think of no better expression to characterize these similarities than "family resemblances," since various similarities which obtain among the members of a family—their build, features, color of eyes, gait, temperament, etc., etc.—overlap and criss-cross in the same way.

This translation differs from Miss Anscombe's (which has been quoted above) in that it makes more explicit the fact that the similarities are similarities among the members of a single family, and are not themselves definitive of what constitutes a *family* resemblance.

¹² Were this aspect of the twofold criterion to be abandoned, and were our use of common names to be solely determined by the existence of overlapping and criss-crossing relations, it is difficult to see how a halt would ever be called to the spread of such names. Robert J. Richman has called attention to the same problem in "Something Common," *Journal of Philosophy*, vol. 59 (1962), pp. 821–830. He speaks of what he calls "the Problem of Wide-Open Texture," and says: "the notion of family resemblances may account for our extending the application of a given general term, but it does not seem to place any limit on this process" (p. 829).

In an article entitled "The Problem of the Model-Language Game in Wittgenstein's Later Philosophy," *Philosophy*, vol. 36 (1961), pp. 333–351, Helen Hervey also calls attention to the fact that "a family is so-called by virtue of its common ancestry" (p. 334). She also mentions (p. 335) what Richman referred to as the problem of "the wide-open texture."

¹³ Ludwig Wittgenstein, *Philosophical Investigations*, §66, p. 31.

¹⁴ Although I have only mentioned the existence of genetic connections among members of a family, I should of course not wish to exclude the effects of habitual association in giving rise to some of the resemblances which Wittgenstein mentions. I have stressed genetic connection only because it is the simplest and most obvious illustration of the point I have wished to make.

It is not my present concern to characterize any feature common to most or all of those activities which we call games, nor would I wish to argue on the analogy of family resemblances that there *must* be any such feature. If the question is to be decided, it must be decided by an attempt to "look and see." However, it is important that we look in the right place and in the right ways if we are looking for a common feature; we should not assume that any feature common to all games must be some manifest characteristic, such as whether they are to be played with a ball or with cards, or how many players there must be in order for the game to be played. If we were to rely exclusively on such features we should, as I have suggested, be apt to link solitaire with fortune-telling, and wrestling matches with fights, rather than (say) linking solitaire with cribbage and wrestling matches with weight-lifting. It is, then, my contention that Wittgenstein's emphasis on directly exhibited resemblances, and his failure to consider other possible similarities, led to a failure on his part to provide an adequate clue as to what—in some cases at least—governs our use of common names.¹⁵

If the foregoing remarks are correct, we are now in a position to see that the radical denigration of generalization concerning the arts, which has come to be almost a hallmark of the writings of those most influenced by recent British philosophy, may involve serious errors, and may not constitute a notable advance.

II

In turning from Wittgenstein's statements concerning family resemblances to the use to which his doctrine has been put by writers on aesthetics, we must first note what these writers are *not* attempting to do. In the first place, they are not seeking to clarify the relationships which exist among the many different senses in which the word "art" is used. Any dictionary offers a variety of such senses (e.g., the art of navigation, art as guile, art as the craft of the artist, etc.), and it is not difficult to find a pattern of family resemblances existing among many of them. However, an analysis of such resemblances, and of their differences, has not, as a matter of fact, been of interest to the writers of the articles with which we are here concerned. In the

second place, these writers have not been primarily interested in analyzing how words such as "work of art" or "artist" or "art" are ordinarily used by those who are neither aestheticians nor art critics; their concern has been with the writings which make up the tradition of "aesthetic theory." In the third place, we must note that the concern of these writers has not been to show that family resemblances do in fact exist among the various arts, or among various works of art; on the contrary, they have used the doctrine of family resemblances in a *negative* fashion. In this, they have of course followed Wittgenstein's own example. The position which they have sought to establish is that traditional aesthetic theory has been mistaken in assuming that there is any essential property or defining characteristic of works of art (or any set of such properties or characteristics); as a consequence, they have contended that most of the questions which have been asked by those engaged in writing on aesthetics are mistaken sorts of questions.

However, as the preceding discussion of Wittgenstein should have served to make clear, one cannot assume that if there is any one characteristic common to all works of art it must consist in some specific, directly exhibited feature. Like the biological connections among those who are connected by family resemblances, or like the intentions on the basis of which we distinguish between fortune-telling and card games, such a characteristic might be a relational attribute, rather than some characteristic at which one could directly point and say: "It is this particular feature of the object which leads me to designate it as a work of art." A relational attribute of the required sort might, for example, only be apprehended if one were to consider specific art objects as having been created by someone for some actual or possible audience.

The suggestion that the essential nature of art is to be found in such a relational attribute is surely not implausible when one recalls some of the many traditional theories of art. For example, art has sometimes been characterized as being one special form of communication or of expression, or as being a special form of wish-fulfillment, or as being a presentation of truth in sensuous form. Such theories do not assume that in each poem, painting,

¹⁵ I do not deny that directly exhibited resemblances often play a part in our use of common names: this is a fact explicitly noted at least as long ago as by Locke. However, similarities in origin, similarities in use, and similarities in intention may also play significant roles. It is such factors that Wittgenstein overlooks in his specific discussions of family resemblances and of games.

play, and sonata there is a specific ingredient which identifies it as a work of art; rather, that which is held to be common to these otherwise diverse objects is a relationship which is assumed to have existed, or is known to have existed, between certain of their characteristics and the activities and the intentions of those who made them.¹⁶

While we may acknowledge that it is difficult to find any set of attributes—whether relational or not—which can serve to characterize the nature of a work of art (and which will not be as vulnerable to criticism as many other such characterizations have been),¹⁷ it is important to note that the difficulties inherent in this task are not really avoided by those who appeal to the notion of family resemblances. As soon as one attempts to elucidate how the term “art” is in fact used in the context of art criticism, most of the same problems which have arisen in the history of aesthetic theory will again make their appearance. In other words, linguistic analysis does not provide a means of escape from the issues which have been of major concern in traditional aesthetics. This fact may be illustrated through examining a portion of one of the articles to which I have already alluded, Paul Ziff’s article entitled “The Task of Defining a Work of Art.”

To explain how the term “a work of art” is used, and to show the difficulties one encounters if one seeks to generalize concerning the arts, Professor Ziff chooses as his starting point one clear-cut example of a work of art and sets out to describe it.

The work he chooses is a painting by Poussin, and his description runs as follows:

Suppose we point to Poussin’s “The Rape of the Sabine Women” as our clearest available case of a work of art. We could describe it by saying, first, that it is a painting. Secondly, it was made, and what is more, made deliberately and self-consciously with obvious skill and care, by Nicolas Poussin. Thirdly, the painter intended it to be displayed in a place where it could be looked at and appreciated, where it could be contemplated and admired. . . . Fourthly, the painting is or was exhibited in a museum gallery where people do contemplate, study, observe, admire, criticize, and discuss it. What I wish to refer to here by speaking of contemplating, studying, and observing a painting, is simply what we may do when we are concerned with a painting like this. For example, when we look at this painting by Poussin, we may attend to its sensuous features, to its “look and feel.” Thus we attend to the play of light and color, to dissonances, contrasts, and harmonies of hues, values, and intensities. We notice patterns and pigmentation, textures, decorations, and embellishments. We may also attend to the structure, design, composition, and organization of the work. Thus we look for unity, and we also look for variety, for balance and movement. We attend to the formal interrelations and cross connexions in the work, to its underlying structure. . . . Fifthly, this work is a representational painting with a definite subject matter; it depicts a certain mythological scene. Sixthly, the painting is an elaborate and certainly complex formal structure. Finally, the painting is a good painting. And this is to say simply that the

¹⁶ I know of no passage in which Wittgenstein takes such a possibility into account. In fact, if the passage from “The Blue Book” to which I have already alluded may be regarded as representative, we may say that Wittgenstein’s view of traditional aesthetic theories was quite without foundation. In that passage he said:

The idea of a general concept being a common property of its particular instances connects up with other primitive, too simple, ideas of the structure of language. It is comparable to the idea that *properties* are *ingredients* of the things which have the properties; e.g., that beauty is an ingredient of all beautiful things as alcohol is of beer and wine, and that we therefore could have pure beauty, unadulterated by anything that is beautiful (p. 17).

I fail to be able to identify any aesthetic theory of which such a statement would be true. It would not, for example, be true of Clive Bell’s doctrine of “significant form,” nor would it presumably be true of G. E. Moore’s view of beauty, since both Bell and Moore hold that beauty depends upon the specific nature of the other qualities which characterize that which is beautiful.

However, it may be objected that when I suggest that what is common to works of art involves reference to “intentions,” I overlook “the intentional fallacy” (see W. K. Wimsatt, Jr., and Monroe C. Beardsley, “The Intentional Fallacy,” *Seewanee Review*, vol. 54 [1946], pp. 468–488). This is not the case. The phrase “the intentional fallacy” originally referred to a particular method of criticism, that is, to a method of interpreting and evaluating given works of art; it was not the aim of Wimsatt and Beardsley to distinguish between art and non-art. These two problems are, I believe, fundamentally different in character. However, I do not feel sure that Professor Beardsley has noted this fact, for in a recent article in which he set out to criticize those who have been influenced by the doctrine of family resemblances he apparently felt himself obliged to define art *solely* in terms of some characteristic in the object itself (see “The Definition of the Arts,” *Journal of Aesthetics and Art Criticism*, vol. 20 [1961], pp. 175–187). Had he been willing to relate this characteristic to the activity and intention of those who make objects having such a characteristic, his discussion would not, I believe, have been susceptible to many of the criticisms leveled against it by Professor Douglas Morgan and Mary Mothersill (*ibid.*, pp. 187–198).

¹⁷ I do not say “all” such definitions, for I think that one can find a number of convergent definitions of art, each of which has considerable merit, though each may differ slightly from the others in its emphasis.

Poussin painting is worth contemplating, studying, and observing in the way I have ever so roughly described.¹⁸

With reference to this description we must first note that it is clearly not meant to be anything like a complete description of the Poussin painting; it is at most a description of those aspects of that painting which are relevant to its being called a work of art. For example, neither the weight of the painting nor its insurable value is mentioned. Thus, whether because of his own preconceptions, or because of our ordinary assumptions concerning how the term "work of art" is to be used, Professor Ziff focuses attention on some aspects of the Poussin painting rather than upon others. In doing so, he is making an implicit appeal to what is at least a minimal aesthetic theory, that is, he is supposing that neither weight nor insurable value need be mentioned when we list the characteristics which lead us to say of a particular piece of painted canvas that it is a work of art. In the second place, we must note that of the seven characteristics which he mentions, not all are treated by Professor Ziff as being independent of one another; nor are all related to one another in identical ways. It will be instructive to note some of the differences among their relationships, since it is precisely here that many of the traditional problems of aesthetic theory once again take their rise.

For example, we are bound to note that Professor Ziff related the seventh characteristic of the Poussin painting to its fourth characteristic: the fact that it is a good painting is, he holds, related to the characteristics which we find that it possesses when we contemplate, observe, and study it. Its goodness, however, is not claimed to be related to its first, third, or fifth characteristics: in other words, Professor Ziff is apparently not claiming that the goodness of this particular work of art depends upon its being a painting rather than being some other sort of work of art which is capable of being contemplated, studied, etc.; nor is he claiming that its goodness is dependent upon the fact that it was intended to be hung in a place where it can be observed and studied; nor upon the fact that it is a representational painting which depicts a mythological scene. If we next turn to the question of how the goodness of this painting is related to the fact that it was "made deliberately

and self-consciously, with obvious skill and care by Nicolas Poussin," Professor Ziff's position is somewhat less explicit, but what he would say is probably quite clear. Suppose that the phrase "obvious skill" were deleted from the description of this characteristic: would the fact that this painting had been deliberately and self-consciously made, and had been made with care (but perhaps not with skill), provide a sufficient basis for predicating goodness of it? I should doubt that Professor Ziff would hold that it would, since many bad paintings may be supposed to have been made deliberately, self-consciously, and with care. Yet, if this is so, how is the maker's skill related to the object's goodness? Perhaps the fact that "obvious skill" is attributed to Poussin is meant to suggest that Poussin intended that "The Rape of the Sabine Women" should possess those qualities which Professor Ziff notes that we find in it when we contemplate, study, and observe it in the way in which he suggests that it should be contemplated. If this is what is suggested by attributing skill to the artist, it is surely clear that Professor Ziff has without argument built an aesthetic theory into his description of the Poussin painting. That theory is implicit both in the characteristics which he chooses as being aesthetically relevant, and in the relations which he holds as obtaining among these characteristics.

If it be doubted that Professor Ziff's description contains at least an implicit aesthetic theory, consider the fact that in one of the passages in which he describes the Poussin painting (but which I did not include in my foreshortened quotation from that description), he speaks of the fact that in contemplating, studying, and observing this painting "we are concerned with both two-dimensional and three-dimensional movements, the balance and opposition, thrust and recoil, of spaces and volumes." Since the goodness of a painting has been said by him to depend upon the qualities which we find in it when we contemplate, study, and observe it, it follows that these features of the Poussin painting contribute to its goodness. And I should suppose that they are also included in what Professor Ziff calls the sixth characteristic of the Poussin painting, namely its "complex formal structure." Thus, presumably, the goodness of a painting does depend, in part at least, upon its formal structure. On the other hand, Professor Ziff

¹⁸ *Op. cit.*, pp. 60-61. It is an interesting problem, but not germane to our present concerns, to consider whether Poussin's painting should be classified as a "mythological" painting, as Professor Ziff describes it, or whether it should be regarded as an historical painting.

never suggests that the goodness of the Poussin painting depends upon the fact that it is a representational painting, and that it has a mythological (or historical) subject matter, rather than some other sort of subject matter. In fact, when he discusses critics such as Kenyon Cox and Royal Cortissoz, Professor Ziff would apparently—and quite properly—wish to separate himself from them, rejecting the view that what makes a painting a good painting has any necessary relation to the fact that it is or is not a representational painting of a certain sort. Thus, Professor Ziff's account of the aesthetically relevant features of the Poussin painting, and his statements concerning the interrelationships among the various features of that painting, define a particular aesthetic position.

The position which I have been attributing to him is one with which I happen to agree. However, that fact is not of any importance in the present discussion. What is important to note is that Professor Ziff's characterization of the Poussin painting contains an implicit theory of the nature of a work of art. According to that theory, the goodness of a painting depends upon its possession of certain objective qualities, that these qualities are (in part at least) elements in its formal structure, and that the artist intended that we should perceive these qualities in contemplating and studying the painting. (Had he not had this intention, would we be able to say that he had made the object self-consciously, deliberately, and with skill?) Further, this implicit theory must be assumed to be a theory which is general in import, and not confined to how we should look at this one painting only. Were this not so, the sort of description of the Poussin painting which was given by Professor Ziff would not have helped to establish a clear-cut case of what is to be designated as a work of art. For example, were someone to describe the same painting in terms of its size, weight, and insurable value (as might be done were it to be moved from museum to museum), we would not thereby learn how the term "work of art" is to be used. In failing to note that his description of the Poussin painting actually did involve a theory of the nature of art, Professor Ziff proceeded to treat that description as if he had

done nothing more than bring forward a list of seven independent characteristics of the painting he was examining. In so doing, he turned the question of whether there are any features common to all works of art into a question of whether one or more of these seven specific indices could be found in all objects to which the term "work of art" is applied. Inevitably, his conclusion was negative, and he therefore held that "no one of the characteristics listed is necessarily a characteristic of a work of art."¹⁹

However, as we have seen, Professor Ziff's description of the Poussin painting was not actually confined to noting the specific qualities which were characteristic of the pictorial surface of that painting; it included references to the relations between these qualities and the aim of Poussin, and references to the ways in which a painting having such qualities is to be contemplated by others. Had he turned his attention to examining these relationships between object, artist, and contemplator, it would assuredly have been more difficult for him to assert that "neither a poem, nor a novel, nor a musical composition can be said to be a work of art in the same sense of the phrase in which a painting or a statue or a vase can be said to be a work of art."²⁰ In fact, had he carefully traced the relationships which he assumed to exist among some of the characteristics of the Poussin painting, he might have found that, contrary to his inclinations, he was well advanced toward putting forward explicit generalizations concerning the arts.

III

While Professor Ziff's argument against generalization depends upon the fact that the various artistic media are significantly different from one another, the possibility of generalizing concerning the arts has also been challenged on historical grounds. It is to Morris Weitz's use of the latter argument that I shall now turn.

In "The Role of Theory in Aesthetics" Professor Weitz places his primary emphasis on the fact that art forms are not static. From this fact he argues that it is futile to attempt to state the conditions which are necessary and sufficient for an

¹⁹ *Ibid.*, p. 64.

²⁰ *Ibid.*, p. 66. For example, Ziff denies that a poem can be said to be "exhibited or displayed." Yet it is surely the case that in printing a poem or in presenting a reading of a poem, the relation between the work and its audience, and the relation between artist, work, and audience, is not wholly dissimilar to that which obtains when an artist exhibits a painting. If this be doubted, consider whether there is not a closer affinity between these two cases than there is between a painter *exhibiting* a painting and a manufacturer *exhibiting* a new line of fountain pens.

object to be a work of art. What he claims is that the concept "art" must be treated as an open concept, since new art forms have developed in the past, and since any art form (such as the novel) may undergo radical transformations from generation to generation. One brief statement from Professor Weitz's article can serve to summarize this view:

What I am arguing, then, is that the very expansive, adventurous character of art, its ever-present changes and novel creations, makes it logically impossible to ensure any set of defining properties. We can, of course, choose to close the concept. But to do this with "art" or "tragedy" or portraiture, etc. is ludicrous since it forecloses the very conditions of creativity in the arts.²¹

Unfortunately, Professor Weitz fails to offer any cogent argument in substantiation of this claim. The lacuna in his discussion is to be found in the fact that the question of whether a particular concept is open or closed (i.e., whether a set of necessary and sufficient conditions can be offered for its use) is not identical with the question of whether future instances to which the very same concept is applied may or may not possess genuinely novel properties. In other words, Professor Weitz has not shown that every novelty in the instances to which we apply a term involves a stretching of the term's connotation.

By way of illustration, consider the classificatory label "representational painting." One can assuredly define this particular form of art without defining it in such a way that it will include only those paintings which depict either a mythological event or a religious scene. Historical paintings, interiors, fête-champêtres, and still life can all count as "representational" according to any adequate definition of this mode of painting, and there is no reason why such a definition could not have been formulated prior to the emergence of any of these novel species of the representational mode. Thus, to define a particular form of art—and to define it truly and accurately—is not necessarily to set one's self in opposition to whatever new creations may arise within that particular form.²²

Consequently, it would be mistaken to suppose that all attempts to state the defining properties of various art forms are prescriptive in character and authoritarian in their effect.

This conclusion is not confined to cases in which an established form of art, such as representational painting, undergoes changes; it can also be shown to be compatible with the fact that radically new art forms arise. For example, if the concept "a work of art" had been carefully defined prior to the invention of cameras, is there any reason to suppose that such a definition would have proved an obstacle to viewing photography or the movies as constituting new art forms? To be sure, one can imagine definitions which might have done so. However, it was not Professor Weitz's aim to show that one or another definition of art had been a poor definition; he wished to establish the general thesis that there was a necessary incompatibility, which he denoted as a logical impossibility, between allowing for novelty and creativity in the arts and stating the defining properties of a work of art. He failed to establish this thesis since he offered no arguments to prove that new sorts of instantiation of a previously defined concept will necessarily involve us in changing the definition of that concept.

To be sure, if neither photography nor the movies had developed along lines which satisfied the same sorts of interest that the other arts satisfied, and if the kinds of standards which were applied in the other arts were not seen to be relevant when applied to photography and to the movies, then the antecedently formulated definition of art would have functioned as a closed concept, and it would have been used to exclude all photographers and all motion-picture makers from the class of those who were to be termed "artists." However, what would the defender of the openness of concepts hold that one should have done under these circumstances? Suppose, for example, that all photographers had in fact been the equivalent of passport photographers, and that they had been motivated by no other interests and controlled by no other standards than those which govern the

²¹ *Op. cit.*, p. 32.

²² To be sure, if no continuing characteristic is to be found, the fact of change will demand that the concept be treated as having been an open one. This was precisely the position taken by Max Black in a discussion of the concept "science." (See "The Definition of Scientific Method," in *Science and Civilization*, edited by Robert C. Stauffer [Madison, Wisconsin, 1949].) Paul Ziff refers to the influence of Professor Black's discussion upon his own views, and the views of Morris Weitz are assuredly similar. However, even if Professor Black's view of the changes in the concept "science" is a correct one (as I should be prepared to think that it may be), it does not follow that the same argument applies in the case of art. Nor does the fact that the meaning of "science" has undergone profound changes in the past imply that further analogous changes will occur in the future.

making of photographs for passports and licenses: would the defender of open concepts be likely to have expanded the concept of what is to count as an art in order to have included photography? The present inclusion of photography among the arts is justified, I should hold, precisely because photography arises out of the same sorts of interest, and can satisfy the same sorts of interest, and our criticism of it employs the same sorts of standards, as is the case with respect to the other arts.

Bearing this in mind, we are in a position to see that still another article which has sometimes been cited by those who argue for the openness of the concept "a work of art" does not justify the conclusions which have been drawn from it. That article is Paul Oskar Kristeller's learned and informative study entitled "The Modern System of the Arts."²³ The way in which Professor Kristeller states the aim of his article suggests that he too would deny that traditional aesthetic theory is capable of formulating adequate generalizations concerning the arts. He states his aim in saying:

The basic notion that the five "major arts" constitute an area all by themselves, clearly separated by common characteristics from the crafts, the sciences and other human activities has been taken for granted by most writers on aesthetics from Kant to the present day. . . .

It is my purpose to show that this system of the five major arts, which underlies all modern aesthetics and is so familiar to us all, is of comparatively recent origin and did not assume definite shape before the eighteenth century, although it had many ingredients which go back to classical, mediaeval, and Renaissance thought.²⁴

However, the fact that *the classification of the arts* has undoubtedly changed during the history of Western thought, does not of itself suggest that *aesthetic theory* must undergo comparable changes. Should this be doubted, one may note that Professor Kristeller's article does not show in what specific ways attempts to classify or systematize the arts are integral to, or are presupposed by, or are consequences of, the formulation of an aesthetic theory.

This is no minor cavil, for if one examines the writers on aesthetics who are currently attacked for their attempts to generalize concerning the nature of art, one finds that they are not (by and large) writers whose discussions are closely allied to the discussions of those with whom Kristeller's article was primarily concerned. Furthermore, it is to be noted that Kristeller did not carry his discussion beyond Kant. This terminal point was justified by him on the ground that the system of the arts has not substantially changed since Kant's time.²⁵ However, when one recalls that Kant's work is generally regarded as standing near the beginning of modern aesthetic theory—and surely not near its end—one has reason to suspect that questions concerning "the system of the arts" and questions concerning aesthetic theory constitute distinct, and probably separate sets of questions. A survey of recent aesthetic theory bears this out. Since the time of Hegel and of Schopenhauer there have been comparatively few influential aesthetic theories which have made the problem of the diversity of art forms, and the classification of these forms, central to their consideration of the nature of art.²⁶ For example, the aesthetic theories of Santayana, Croce, Alexander, Dewey, Prall, or Collingwood cannot be said to have been dependent upon any particular systematic classification of the arts. In so far as these theories may be taken as representative of attempts to generalize concerning the arts, it is strange that current attacks on traditional aesthetics should have supposed that any special measure of support was to be derived from Kristeller's article.

Should one wish to understand why current discussions have overlooked the gap between an article such as Kristeller's and the lessons ostensibly derived from it, an explanation might be found in the lack of concern evinced by contemporary analytic philosophers for the traditional problems of aesthetic theory. For example, one looks in vain in the Elton volume for a careful appraisal of the relations between aesthetic theory and art criticism, and how the functions of each might differ from the functions of the other. A

²³ *Journal of the History of Ideas*, vol. 12 (1951), pp. 496-527, and vol. 13 (1952), pp. 17-46. This study has been cited by both Elton (*op. cit.*, p. 2) and Kennick (*op. cit.*, p. 320) in substantiation of their views.

²⁴ *Op. cit.*, vol. 12, p. 497.

²⁵ *Op. cit.*, vol. 13, p. 43; also, pp. 4 ff.

²⁶ One exception is to be found in T. M. Greene: *The Arts and the Art of Criticism* (Princeton, 1940). This work is cited by Kristeller, and is one of the only two which he cites in support of the view that the system of the arts has not changed since Kant's day (*op. cit.*, vol. 12, p. 497, n. 4). The other work cited by him is Paul Franke's *System der Kunstwissenschaft* (Brünn/Leipzig, 1938), which also offers a classification of the arts, but only within a framework of aesthetic theory which could easily embrace whatever historical changes the arts undergo.

striking example of the failure to consider this sort of problem is also to be found in John Wisdom's often cited dicta concerning "the dullness" of aesthetic theory.²⁷ In examining his views one finds that the books on art which Wisdom finds *not* to be dull are books such as Edmund Wilson's *Axel's Castle*, in which a critic "brings out features of the art he writes about, or better, brings home the character of what he writes about."²⁸ In short, it is not theory—it is not aesthetic theory at all—that Wisdom is seeking: he happens to be interested in criticism.

I do not wish to be taken as denying the importance of criticism, nor as belittling the contribution which a thorough acquaintance with the practice of criticism in all of the arts may make to general aesthetic theory. However, it is important to note that the work of any critic presupposes at least an implicit aesthetic theory, which—as critic—it is not his aim to establish or, in general, to defend. This fact can only be overlooked by those who confine themselves to a narrow range of criticism: for example, to the criticism appearing in our own time in those journals which are read by those with whom we have intellectual, political, and social affinities. When we do not so confine ourselves, we rapidly discover that there is, and has been, an enormous variety in criticism, and that this variety represents (in part at least) the effect of differing aesthetic preconceptions. To evaluate criticism itself we must, then, sometimes undertake to evaluate these preconceptions. In short, we must do aesthetics ourselves.

The Johns Hopkins University

However, for many of the critics of traditional aesthetics this is an option which does not appeal. If I am not mistaken, it is not difficult to see why this should have come to be so. In the first place, it has come to be one of the marks of contemporary analytic philosophy to hold that philosophic problems are problems which cannot be solved by appeals to matters of fact. Thus, to choose but a single instance, questions of the relations between aesthetic perception and other instances of perceiving—for example, questions concerning psychical distance, or empathic perception, or the role of form in aesthetic perception—are not considered to be questions with which a philosopher ought to try to deal. In the second place, the task of the philosopher has come to be seen as consisting largely of the unsnarling of tangles into which others have gotten themselves. As a consequence, the attempt to find a synoptic interpretation of some broad range of facts—an attempt which has in the past been regarded as one of the major tasks of a philosopher—has either been denigrated or totally overlooked.²⁹ Therefore, problems such as the claims of the arts to render a true account of human character and destiny, or questions concerning the relations between aesthetic goodness and standards of greatness in art, or an estimate of the significance of variability in aesthetic judgments, are not presently fashionable. And it must be admitted that if philosophers wish not to have to face either factual problems or synoptic tasks, these are indeed questions which are more comfortably avoided than pursued.

²⁷ See "Things and Persons," *Proceedings of the Aristotelian Society, Supplementary Volume XXII* (1948), pp. 207–210.

²⁸ *Ibid.*, p. 209.

²⁹ For example, W. B. Gallie's "The Function of Philosophical Aesthetics," in the Elton volume, argues for "a journeyman's aesthetics," which will take up individual problems, one by one, these problems being of the sort which arise when a critic or poet gets into a muddle about terms such as "abstraction" or "imagination." For this purpose the tools of the philosopher are taken to be the tools of logical analysis (*op. cit.*, p. 35); a concern with the history of the arts, with psychology, or a direct and wide-ranging experience of the arts seems not to be presupposed.

A second example of the limitations imposed upon aesthetics by contemporary linguistic analysis is to be found in Professor Weitz's article. He states that "the root problem of philosophy itself is to explain the relation between the employment of certain kinds of concepts and the conditions under which they can be correctly applied" (*op. cit.*, p. 30).

VI. UNFALSIFIABILITY AND THE USES OF RELIGIOUS LANGUAGE

ALASTAIR McKINNON

I

JUST thirty years ago the logical positivists condemned religious statements as meaningless on the ground that they were not subject to any kind of empirical verification. Since then the charge has been altered to read that they are unfalsifiable. But apart from this there has been little progress. Indeed, many philosophers now dismiss this charge as uninteresting and unimportant. This is not without some justification. The charge proved much less helpful than many had expected. It begged the question it presumed to settle and it served to obscure rather than to reveal the real nature and logic of the discourse it was used to condemn. But despite all this it seems that there is still some juice in this particular lemon. In any event, I propose to give it at least one more squeeze.

The charge is familiar and quite straightforward: Religious utterances are unfalsifiable and this despite the fact that they are apparently intended as assertions. The point can be seen by putting such utterances beside ordinary empirical assertions. The latter are defeated or falsified by a single counter-instance. "All swans are white" is routed by the appearance of a single swan of any other color and one who asserts this statement must surrender it in the face of conflicting evidence. But, the charge runs, religious utterances behave quite differently. Though they look like factual claims and are frequently advanced as such, nothing is ever allowed to tell finally against them. Here nothing is permitted to stand as a counter-instance. They are unfalsifiable and, because of their appearance, dishonest. They masquerade as factual statements and claim all the benefits thereof but refuse the test of falsification proper to all such statements. They are empty or vacuous, mere pseudo-assertions which in fact assert nothing at all.

This charge can be expressed in terms of the notion of belief. In its most familiar sense, belief is the acceptance of a factual or at least quasi-

factual proposition: it is the claim that there are no counter-instances. Thus conceived, belief is a relatively simple and straightforward notion. But, the charge goes, religious belief refuses to abide by this simple rule. It is notorious that the believer refuses to surrender his claim even when there is apparently conclusive evidence against it. Believers are in fact like Job: they cling to their beliefs even when all that seems to have prompted them has been destroyed. Now we know what it means to believe in the ordinary sense but what can it mean to believe in the face of apparently clear counter-instances? In fact, can it mean anything at all?

Recent discussions of this and related charges have employed as examples "God is love" and, less explicitly, "God exists." The former is in order and will serve as our first example. But the latter is open to grave suspicion. Philosophers do indeed consider this claim and they repeatedly ascribe it to believers. But the truth is that most believers do not use this form. Neither, as we shall see, should they do so. Instead, with the Apostles' Creed, they say, "I believe in God. . . ." We shall therefore take this properly religious utterance as our second example. The difference between the two is very considerable. As we shall see, the latter is quite distinct from the simple theistic thesis which philosophers traditionally ascribe to religious believers. This itself is a matter of great importance. It may be interesting to inquire concerning that thesis but it is obviously of much greater interest to explore the things which believers actually say.

Our discussion of these examples is marked by a steady emphasis upon the various contexts in which they are actually employed. Philosophers have tended to treat utterances as though they had meaning in and of themselves and this without reference to context or the purpose they were intended to serve. With one or two exceptions this has been true even of those who sought to defend religious utterances against the charge before us. This is a serious error for the simple reason that here at least such matters are of the very greatest

importance. Religious utterances occur in specific contexts and their meaning can be discovered only by a careful examination of these contexts. This is the chief task before us. Our immediate concern is to distinguish the three most important contexts in which these examples are employed and hence their three principal meanings or uses. When these have been enumerated it will be clear that it makes no sense to ask about the unfalsifiability of these utterances as such. Indeed, it will be seen that the question really is which of these uses are unfalsifiable and which are not.

In order to set forth these uses more clearly we shall begin with a model having the same complex relation to scientific activity as our own examples have to religious belief and practice. The philosophical critics of science have provided a number of such claims but we shall content ourselves with the simple and obvious: "The world has an order." There is good reason for beginning in this way. Though the scientists have not explicitly answered their critics their actual use of this and similar statements provides the basis for an answer to the charge before us. This charge rests ultimately upon certain traditional and restrictive theories about the way in which words and, derivatively, utterances can function. It seems therefore best to begin with a model from another realm, a model which actually functions in a way not allowed by these theories. When its behavior has been charted we can then proceed to describe the various uses of our examples and hence of religious utterances as such.

For both the model and our examples we shall distinguish the three most important typical uses. These we shall describe as the *assertional*, the *self-instructional* and, for want of a better term, the *ontological-linguistic* use (or sense). These distinctions are closely linked with two different uses of the key or operative concept which we shall describe as the *determinate* and the *heuristic* use (or sense) and indicated as, for example, love(m) or (n) or (o) and love(X), respectively. The ordinary letters, chosen from the middle of the alphabet, stand for successive determinations of the concept while the capitalized X, borrowed from simple algebra, stands for a definite but nevertheless as yet unknown value. It goes without saying that in all cases both the various sense of the term and the different uses of the utterance reflect the intention of the speaker and, behind this, the context of his utterance.

The course of our argument is both obvious and simple. It begins by discriminating the three

distinct senses in which the scientist uses our model. It then considers the two examples in turn, in each case distinguishing the believer's corresponding uses of these utterances and, further, describing these three uses in some detail. Finally, upon the basis of this analysis, it asks which of these uses, if any, are unfalsifiable and, if so, whether this is a serious defect.

II

Though philosophers have tended to concentrate upon the terms of statements their meaning is in fact largely determined by the context in which they occur. This is certainly true of the scientist's "The world has an order," an utterance with three typical contexts and, corresponding thereto, three distinct meanings or uses.

Imagine a scientist who has finally discovered a causal link between *A* and *B*. He has unearthed a connection and later will spell it out in detail. But for the moment his concern is simply to report or announce his discovery. He can do this in a number of ways: "There was a connection after all," "Yes, I have discovered the law," or, in a more expansive mood, "The world has an order." He means, of course, a particular connection, a particular law and, equally, a particular order.

This use of "order" is quite clear. The scientist has discovered a particular order or uniformity and wishes to report this fact. Hence "order" is simply shorthand for a particular or determinate order, an order which he could produce or specify upon demand. This is the determinate sense or use of the term and can be marked as order(m).

Leaving aside the relatively complicated question of the existence of such orders (Does anyone now wish to assert the existence of Newtonian gravity?), the nature of this utterance seems quite clear. It is intended as a factual claim on all fours with "There is a cat in the next room," "There is a tree on the lower campus," etc. It asserts a state of affairs (that *A* is the cause of *B*) and can be falsified by a single counter-instance (a case where *B* is not preceded by *A*). This is the assertional use of the utterance and can be expressed as "The world has order(m)."

But this same utterance also appears in another and perhaps more revealing context. Imagine that our scientist has failed to discover the suspected causal connection; imagine even that he has encountered evidence which tells decisively against conceptions he has long held. He could, of course, give up in despair; he could decide that there was

no hope of making sense of all the evidence before him. Alternatively, he could recall his scientific commitment to see every single part as coherent with the whole. In that event he might well steady himself with the remark, "Everything must fit together," "The universe is a single whole," or, equally appropriate, "The world has an order."

It is of course clear that in this case the scientist is no longer using the word "order" as shorthand for some determinate or specifiable regularity. Again, it is equally clear that he does not intend his utterance in anything like its earlier sense. Indeed, in the present context, he cannot intend a claim about the world at all. He is rather telling himself to treat every event as part of the world; to see all phenomena, however strange, as falling within its order. Of course, he is not enjoining himself to force this event into some conception he already has. Instead, he is instructing himself to get on with the job of formulating a conception having room for this event. Again, he is using "order" not to stand for this or that particular order but simply for order as such. He is using it in the sense of "form," "shape," or "way." He is using it in its heuristic sense and this we mark as order(X).

It is important not to be misled by the linguistic form of "The world has an order." Here it is not a claim about the world but rather a pledge or vow which the scientist is taking. It is his self-command to get on with the job, to continue in the scientific enterprise. This is the self-instructional use and can be expressed as "Treat every single event as relevant to the proper determination of the world's real order(X)" or, more briefly, "Treat all events as part of the world's order(X)." Alternatively, it can be expressed as "Understand 'order' in an heuristic sense." This is, significantly, an equivalent form of self-instruction.

But there is another and very different context in which the scientist would use this utterance. Imagine that he is attempting to meet the philosophers' challenge to state and defend the so-called assumptions of his discipline. Imagine, for that matter, that he is trying to justify the procedure already described. Such situations are rare but the use they engender is interesting and highly significant.

It is important to distinguish at the outset between the answer which the scientist actually gives and that which the philosopher expects and requires of him. Since the time of Hume philosophers have been prone to assume that science can be justified only by showing that this is indeed a

certain kind of world. They have required a demonstration or, at the very least, a reasonable assurance, that the world was a certain way: for example, that it had an objective pattern, that it was sufficiently orderly, that it had the requisite degree of simplicity, that it was suitably geared to the human mind, etc. They have required an assurance concerning what was essentially a contingent matter of fact. And the scientist who accepts this challenge on its own terms commits himself to giving such an assurance. He commits himself to saying something essentially similar to that which he says in the first situation; similar, but, of course, much more general. Such a scientist, if he used our utterance, would intend it in what might be called its metaphysical sense. Of course it is precisely this use which the scientist seeks and must seek to avoid.

There is of course a standard and telling objection against the justification of science by the results of any empirical investigation. Since these results come from such investigation they cannot stand as its foundation or justification. Philosophers who continue to demand a metaphysical justification for science might well recall Hume's insight that such justification is logically impossible.

But if this standard objection shows the impossibility of such a justification there is another objection which points to the true one and, further, the one which the scientist intends. The underlying difficulty with the metaphysical one is that it assumes an essentially pretentious and unrealistic account of the insight or understanding afforded by science and, as a result, accounts for the intelligibility of the world in terms of what is essentially a happy accident. In fact, science does not require that the world have this shape or that. It requires only that the world have some shape or other. And this it cannot lack. The intelligibility of the world, at least as discovered by science, so far from being a happy accident, is rather a necessary consequence of its being anything at all.

We can perhaps now better understand the sense in which the scientist might here intend "The world has an order." Clearly he cannot employ "order" in a determinate sense. Nor can he in this situation ascribe some determinate order to the universe. He cannot now make any empirical or even metaphysical assertions about the world. But it would be mistaken to conclude that he cannot employ the term "order" at all or that he cannot significantly apply it to the world. Whether such

use and application is possible is the crucial question and it is not to be settled by mere *a priori* prejudice or the acceptance of familiar theories about the way in which language ought to function.

It is widely supposed that we can significantly assert a predicate only if we can in some sense specify that predicate; that, for example, we can say "The world has an order" only if we are able to spell out what we mean by "order." But words function in many ways and this condition, though ordinarily satisfied, is not always necessary. As in the previous case, it is possible to employ this term in an heuristic sense. And this is precisely how the scientist now intends it. It is not shorthand for some determinate order but rather a synonym for "form," "shape," or "way," a sense we can mark as order(X). When he so uses "order" to justify his work he is employing our model utterance in its ontological-linguistic sense. This use we can express as "The world has order(X)."

The same assumptions which prompt us to suppose that predicates must be specifiable also prompt us to think that their use in a determinate sense is a condition of our making a real statement or claim about the world. But again this is not true. "Order" need not be given a determinate use; it need not be intended in the sense of order(m) or (n) or (o). It can instead be employed in an heuristic sense; it can be intended in the sense of "the way the world is, however that may be." And, so used, it can be part of a genuine factual claim about the world. Further, because the world could not conceivably lack order in this sense, this particular claim possesses an interesting and important feature: unlike most factual claims it is both true and necessarily true.

This takes care of the first or "ontological" part of our description but it is important to recognize the second and, particularly, to see their connection. The importance and role of the "linguistic" aspect can be put quite simply. Knowing that the world has and must have order in the heuristic sense is not at all like knowing one or more of the ordinary empirical details of the world. It follows simply and directly from understanding the scientists' heuristic use of this term and its necessary connection with the world. "Order" is not simply what we have learned to call an open-textured concept. It is not simply that its lines are somewhat shifting and blurred. It is rather that what is ordinarily called the content or meaning of this concept is determined finally and entirely by

the actual character of the world: its meaning is simply a copy of the actual shape of the world, whatever that should prove to be. In short, the utterance, and the truth of the utterance, follows from the heuristic use of this term. But, it must be added, this utterance is not in any sense a merely linguistic truth. The fact it asserts is no less a fact, and no less important for the fact, that it follows and follows necessarily from a certain use of a word. Science does not rest upon some merely contingent matter of fact. Its real and only necessary foundation is something which could not be other than it is: it is simply the presence of order in a sense which the world could not conceivably lack. It is to underline and protect this truth that we have described this as the ontological-linguistic use.

One further remark must be made in this connection. Recently we have been encouraged to draw a sharp distinction between the empirical and the linguistic but is not this distinction largely illusory? Perhaps it is the fiction of a mind enchanted by the hope of breaking out of the prison of its own language. Perhaps we have been led astray by a now famous aphorism; perhaps we now need to be freed of one of Wittgenstein's own bewitchments. Perhaps we should say, "A cow is a cow is a cow. . . ." Or, better, "A cow is what it is and we call it a cow."

Our concern is with the charge of unfalsifiability as directed against religious utterances but it may be helpful to ask briefly how far this charge may be properly urged against the scientist's various uses of his utterance. It is of course clear that this charge is scarcely appropriate to its second use. Pieces of self-instruction may be sound or ill-advised; they may be difficult or even incapable of implementation. But they cannot be true or false and the demand that they be falsifiable is simply confused and mistaken. But matters are quite different in the case of the first use. Though this one can scarcely be true in the strict sense it nevertheless aims after truth. It is therefore falsifiable and, it might be added, repeatedly falsified. But, and this is a most important point, because of his approach to the subject, because he is constantly putting his material together in a new way, the scientist repeatedly furnishes a new determinate meaning for "order," a meaning which he urges in place of the one he has been forced to discard. Order(m) gives way to order(n) and this in turn to order(o) and so on. Hence though evidence does tell and tell decisively against what he intends to assert, it does

not and cannot tell finally against that utterance itself. Utterances like "The world has order(m)" are repeatedly defeated, but "The world has an order," just because it is a loose and elliptical form, is never subject to final defeat. The case of the third sense is of course quite different. Here the scientist is not alleging that the world is some particular way; he is not making a claim which might either be supported by evidence or refuted by further evidence. Rather he is saying that the world has order in the heuristic sense and that is a claim which, because the term is so used, could not possibly be false. The world might lack any particular order; it might even lack the order it presently has; but it does not and cannot lack order as such. What the scientist intends in this case is both true and necessarily true. It is therefore entirely mistaken to think that it is falsifiable or to suppose that it should be.

III

No doubt there are some important differences between the scientific utterance just considered and the religious examples we are now about to discuss. Nevertheless none of these differences are, I think, relevant to the question before us. In any event the believer does use "God is love" in three exactly parallel situations and, hence, in three precisely corresponding senses.

Of the various uses of "God is love" the first and most common is the assertional one. It is to make a factual claim about God. The believer has recognized some event as an expression of God's love or, alternatively, has come through this event to some new and perhaps fuller understanding of this concept. In any event he now intends to predicate this quality or character of God. Of course, he may feel that his conception of love is not entirely adequate; he may even hold that it does not do justice to the real nature of God's love. Nevertheless these sophisticated and proper doubts do not obtrude upon his normal use of this utterance. In the situation or context he uses "love" in a determinate sense and he consciously intends to assert this of God. His meaning can then be expressed as "God is love(m)."

Though this use of the utterance is perhaps not the most decisive or distinctive one it is nevertheless of very great importance. Such uses play a prominent part in the life of worship and they constitute an important part of the beliefs of the religious man. It is, in fact, such claims which

constitute what is usually called the content of belief, and it is in such cases that we speak of someone as *believing that*.

We have described this as the assertional use of the utterance and the believer clearly intends it in this way. Nevertheless this is not strictly correct. Because "love" may stand for a host of different determinate values, the utterance as here used, is not so much an assertion as a blank for a variety of possible assertions. This is central for the question of the unfalsifiability of this use.

A second and more decisive use of this utterance occurs in situations which are not uncommon in the life of the believer. A striking example is that of the young man who, having known only good fortune and happiness, now learns that he is dying of cancer. It is merely a prejudice to insist that in this situation such a person must intend "God is love" simply as an habitual and purely automatic response. So too it is prejudicial to suppose that he must recognize his situation as an expression of God's love or that he intends his utterance as an assertion. In any event, linguistic form is a notoriously unreliable guide to actual use. In fact there is no reason to suppose that the believer who uses this utterance in such situations actually intends a factual claim. Rather, I suggest, he is doing something quite different and, finally, much more important. He is engaging in self-instruction or, perhaps, self-pledging. He is enjoining himself to see this presently incomprehensible event as revealing, however darkly, something more of the nature of God's love. He is using "love" in an heuristic sense. By the same token he is using the entire utterance in what we have called the self-instructional sense. His meaning is then best expressed as "I must treat all events as pointing to God's love(X)."

This use is typical of the religious believer and indicative of a very important element in the religious life. Religion does not consist entirely in the holding of certain beliefs; equally important is the acceptance of a discipline to act in a certain way. And it is this which the believer undertakes when, in such situations, he deliberately says "God is love." He is steadying himself, binding himself to respond in a certain way. This may still be described as belief but it should be qualified as *belief in* rather than *belief that*.

As with our model, the third use of this utterance occurs in the face of criticism and, perhaps, self-doubt. The sceptical critic sees the young man's plight as clear proof that God is not love and as

decisive evidence against the believer's claim. And perhaps the young man himself begins to doubt. Such situations, however rare, bring out the distinctive character of the third use of this utterance.

It might be thought that the believer is here claiming to see how everything, including his present misfortune, is indeed an instance of divine love. This would make his utterance a straightforwardly factual claim corresponding to the metaphysical use of the scientific model. But this use is open to the same objections as in the previous case. In any event the young man does not know the meaning or content of love in a way which would permit him to see how his condition squares with God's love. He is not in a position to see his misfortune as an expression of that love. Indeed, if he could so see it he would not be in his present difficulty.

There is, of course, another obvious interpretation of this utterance. It might be argued that the young man is using "love" analogically. On this view he would take the highest conception of love which humans can know and apply this analogically to God. Now no doubt such terms are sometimes so used; no doubt many have been taught so to use them. I want only to suggest that in such situations they are in fact used in another and very different sense. That use has already been seen in the case of our scientific model. The scientist can say "The world has an order" even when he is unable to specify that order. And he can know that his utterance is true even when, in the ordinary sense, he does not know the meaning of "order." This is because for the scientist order is simply a copy of the way the world is.

The believer's use of his utterance in this situation is exactly parallel. He does not know the meaning of love in the ordinary sense; he cannot produce an account of love of which his present situation is clearly an instance. Nevertheless he does know that for him, as a believer, it is God's action alone which finally determines the true meaning of this conception. He does not yet know its full meaning but he does know the rules for its determination. Because of this he can assert and know as true the claim "God is love" in a sense best expressed by the form "God is love(X)." In our own terms, he can and does use "love" in an heuristic sense and, equally, he can and does use the entire utterance in an ontological-linguistic sense.

A brief word in explanation of this description.

The believer is indeed making a factual claim about God but one which, unlike ordinary factual claims, could not be false. The scientist's use of "order" is such that the world could not lack it. Similarly, at least this point or extremity, the believer's use of "love" is such that God could not fail to show or express it. Of course, this does not bind God's action any more than "The world has an order" places a restriction upon nature. It is simply a way of saying that for the believer "love" is finally defined from the outside.

IV

Our second example "I believe in God" occurs in the same three contexts and has the same three principle uses. However it does appear to differ from the preceding utterances in two minor aspects and it is necessary at least to mention these.

"I believe in God" consists of a declarative and an assertive aspect which might be marked as "I believe . . ." and " . . . that God exists." Or, rather, it consists of a declarative form and an assertive element. This cursory analysis suggests the way in which the first aspect influences the second and, depending upon the context, makes the latter something quite distinct from the traditional "God exists." It also directs attention to the important fact that religious utterances are something more than the assertion of fact conceived as such. This is perhaps most obvious in the case of the second use where the declarative aspect is a reliable index of the actual use. In the case of the first and, for quite different reasons, the third it is the assertive aspect which is of primary interest. In each case we shall attempt to interpret this element in the light of the relevant context.

There is one other difference between "I believe in God" and the preceding example. While the latter was concerned with the nature of God the former appears to touch only upon his existence. But, as we shall see, the two questions are much less distinct than the tradition has generally supposed. In fact, generally speaking, the question whether God exists is unanswerable apart from some assumptions concerning his nature. The two questions are in fact interrelated and, of more immediate interest, the difference between these two utterances is apparent rather than real.

The believer's normal and typical use of "I believe in God" is to make the factual claim that God, as the believer now conceives him, does indeed exist. The God in question may be that of the

believer's childhood or that of some apparently sophisticated scientist; in either case he has some determinate conception in mind and intends to assert that something exists corresponding to this conception. His meaning is then best conveyed as "God(m) exists." And this is of course the assertional use of this utterance.

There is no doubt that, as a matter of fact, such claims and the belief expressed therein play an important and even indispensable part in the life and thought of the believer. But it would be entirely wrong to suppose that this is the whole story of his belief. Belief may well begin here but if it never moves beyond this point it can scarcely be worthy of that name.

Of course this utterance is confronted with the same difficulties and possessed of the same features as our earlier example. The believer wishes to assert the existence of God(m) and yet, if he is even moderately sophisticated, he understands that this conception is inadequate and that his claim therefore could not be strictly true. Again, though intended as an assertion or claim, his utterance, as ordinarily expressed, is not so much a factual claim as a blank for a set of such claims for all of which it is equally appropriate.

As suggested by its alleged resemblance to the previous example, the second use of "I believe in God" is appropriate in situations such as that of the young cancer victim. Just as this man is not now in a position to see how God is love neither is he able to see how his plight may be squared with God's existence. And, significantly, he does not say, "God exists." Instead, he says "I believe in God." And in so saying he is pledging himself to undertake to see his plight as, together with the rest of experience, pointing to the true nature of the God who exists. He is using the term "God" in an heuristic sense. This time he is saying "Treat all events as pointing to the real nature of God(X)."

This instruction or pledge can of course be expressed in a different way. It may be put as "Understand 'God' in an heuristic sense" or, alternatively, "Allow the content of 'God' to be determined by events." This is simply to put the same resolution in terms of the meaning of words.

This use of the utterance is at least as common as the first; in actual practice perhaps the two are set in a kind of balance within the life and thought of the believer. But whatever the relative frequency of this use it is again clear that religious belief is something more than and different from commitment to certain merely factual claims. It is also

belief in. To believe in this sense is to interpret events in a certain way. It is to commit oneself to a certain course of action. In short, belief, at least in this sense, does make a difference.

To grasp the third use of this utterance we have only to imagine our young victim surrounded by his critics and perhaps beset by his doubts. The critic wants to be shown how this tragedy can be reconciled with the existence of God. Or, rather, he wants a conception of God with which this could be reconciled. And, one might suppose, nothing less can assure even the believer. But, of course, he cannot supply this conception and just herein lies his problem.

The plight of the believer is of course precisely similar to that of the scientist who is required to justify his practice. And, not surprisingly, he responds in a precisely similar way. For the solution of his problem does not consist in claiming to see what plainly he does not see; it does not lie in pretending to have a meaning for "God" which will cover his tragedy. It consists rather in adopting heuristic sense for the word "God," in seeing this word as a marker for something like "the Ultimate Reality, whatever that should prove to be." But there can be no question of the existence of God thus conceived. That there is such a God is a necessary and indubitable truth, a truth which is distorted in the traditional "God exists" but which can perhaps be expressed in the form "God(X) exists." This formulation in any event makes it clear that one is not asserting the contingent existence of a Being but rather the necessary existence of Reality or Being as such.

The contrast between this position and the traditional one might be put in the following simple way. Philosophers who have considered the questions of the grounds of religious belief have treated the religious life as if it were something like a wild animal hunt; a hunt upon which one could not reasonably embark without first having a guarantee or at least a very strong assurance that there was indeed an animal in one's assigned territory. In fact, however, that life is much more like the game of identifying the largest elephant in a herd already before us. Provided only there is such a herd, the expression "the largest elephant in the herd" must necessarily have some counterpart in reality. Again, provided only that there is something which is real (and it is surely philosophical in the bad sense to quibble on this point) the expression "Ultimate Reality" or "God" thus understood must also have an actual counterpart.

Big game safaris may indeed require advance scouts but it is entirely possible that what religion requires is simply the accurate description of the various ways in which its utterances are actually used.

Earlier we said that the question of the existence of God was generally inseparable from his nature. And certainly this is so in the first use. But it is not so in the present one. The believer who intends "God(X) exists" is not laying claim to a knowledge of the nature of God; indeed, he is insisting upon the inadequacy of such knowledge as he already has. Nevertheless, he is saying that God, thus conceived, does and must exist. And this insight is the real foundation of religion in quite the same sense as the corresponding conception of order is the real basis of science.

We have alleged that religious utterances are actually used in three quite separate and distinct senses. Before proceeding it may be well to offer some evidence for this claim, at least in respect to the less obvious cases. This is perhaps not necessary for the first use. Despite its difficulties everyone recognizes this as a common and familiar use of such utterances. Nor is it necessary to argue at length for the second. This use reflects a central aspect of the religious life and one which, if generally neglected in philosophical accounts of belief, is nevertheless acknowledged by all who have experience of such life and belief. Concerning the third, however, there may be some doubt. This is not simply because traditional theories of language tell against the possibility of such use. Equally important is the realization that religious belief is in some important sense focused upon a determinate conception of God. This is a fact which is widely known and to which even specifically devotional writings abundantly testify. But religion has another quite different and equally important aspect. Though worship is indeed directed upon a determinate conception of God it nevertheless acquires its distinctive quality and tone from the sure knowledge that God is something other than and different from any human conception of him. The worshipper necessarily "works" with his own conception of God but he is capable of real worship only to the extent that he realizes that God necessarily surpasses both his own and any possible human conception. Indeed, it is just this realization which marks the line between real worship and mere idolatry. Even orthodox theology, itself perhaps tending in the opposite direction, makes this point with its insistence upon

the infinity and essential incomprehensibility of God. This is therefore an essential and acknowledged aspect of the properly religious life. And it is, of course, this aspect which underlies and is reflected in the third use of our examples.

One further point. Whether these uses will actually pass the unfalsifiability test is a question yet to be decided. In any event the reader may rest assured that they have not been selected with this end in view. Thus far our only concern has been to report and describe the more important and distinctive ways in which believers actually use religious language. We have attempted a faithful report and if it seems forced or strained this perhaps may be due to a long neglect of the meanings which believers actually intend in favor of those which philosophers have repeatedly attributed to them.

V

In the preceding analysis we have spoken as though any actual utterance was a pure and clear instance of one or other of the uses described. In fact, of course, there may be situations in which the believer intends two or perhaps even all three of these senses. And no doubt in many cases he could not say precisely what he has in mind. But however that may be, it is now clear that the question of the unfalsifiability of religious beliefs is not so simple as has been widely assumed. The fact is that these beliefs are complex and varied and there is and can be no simple answer to this charge. It must rather be an answer in terms of each of the different uses of these beliefs.

The self-instructional use of religious beliefs is central to the religious life even in its narrowly religious aspects. But such beliefs, when so used, are not subject to falsification. A piece of self-instruction may be good or bad; it may be wise or ill-considered. But it cannot be true or false. It is not even the sort of thing which might be either true or false. Hence the demand that such uses be subject to falsification is simply mistaken. It betrays a confusion concerning the character of this use and perhaps even a failure to note its presence and importance.

The assertional use of such beliefs states or expresses a factual claim and is therefore properly subject to the falsifiability test. But the application of this test is not so simple as one might imagine. As already suggested, utterances when so used are not so much assertions as a blank or shorthand for

a variety of assertions, a variety made possible by the fact that the key term in such utterances is really a variable for a number of different determinate conceptions. And because this use travels together with the self-instructional one, because the two are constantly set against each other, one determinate conception constantly replaces another and new factual claims are urged to replace those already defeated. Hence while evidence can and does tell against any instance of this use it cannot tell against the use itself. The believer's factual claims are frequently defeated but the form in which these claims are made need not itself ever be subject to final defeat.

The ontological-linguistic use of religious beliefs shows important differences from both the preceding ones. Because religious utterances, so used, assert a factual claim it might be supposed that they were proper candidates for the falsifiability test. But such utterances are factual with a difference. Though factual in a perfectly proper sense they are also, as we have insisted, both true and necessarily true. As here used, they could not be false. There is nothing that could conceivably tell against them. Hence the protest that such claims or beliefs are unfalsifiable is entirely beside the point. They simply are so and this is neither a defect nor a virtue. It does, however, follow from their peculiar nature and is connected with their function as the real foundation of belief.

VI

There are no beliefs or utterances in the abstract. These always occur in a context and have a sense

or use connected therewith. The question therefore is not whether an utterance as such is falsifiable but rather which of its uses are falsifiable and which are not. These comments have sought only to answer this question in respect to some of the more interesting ones.

If this analysis is correct the critic's annoyance with the believer's apparent imperturbability is less well founded than he supposes. It is of course true that the believer's statements look like simple and straightforward claims; is is even true that many believers conceive their beliefs in this way. But the truth is that these are much more complex than the critic supposes. And his annoyance is due, in some measure at least, to his failure to note this complexity. But the critic has another and perhaps deeper grievance. He is genuinely distressed by the fact that even in the face of apparently decisive evidence religious persons are hesitant to surrender their belief. Here two facts should be noted. Believers do in fact give up their beliefs. Second, they do not do so as readily here as in many other matters. The preceding account of the uses of religious beliefs, I submit, explains and connects these facts.

Finally, a point about the word "belief." One may *believe* a religious claim in the first use. One may *believe in* it in the second. But "belief" is not sufficiently strong for the third. Beliefs cannot be based upon what is itself simply another belief. How the believer's relation to this third use should be described I leave to the judgment of the reader. Throughout this paper I have presented it simply as a matter of using certain words in their distinctive but nevertheless wholly proper sense.

VII. FAMILY RESEMBLANCE PREDICATES

KEITH CAMPBELL

AS a step toward showing how it can be proper to employ the same predicate in describing many different individuals (solving the problem of universals in predicative position), logicians formerly used to place all proper predicates in one or other of two classes:

- (i) The *simple*: those which applied where something is common to all individuals truly describable by the predicate in question, the exact nature of this something in common remaining controversial.
- (ii) The *complex* or *definable*: those which are a presence-function of members of class (i). *F* is a presence-function of *G* and *H* if and only if every sentence ascribing *F* to an individual is a truth-function of ascriptions of *G* and *H* to that individual.

Let the doctrine that every proper predicate belongs to one of these two classes be dubbed The Traditional Dichotomy.¹ This Dichotomy is to be regarded as one feature of an abstract model for European languages, whereby the propriety of these languages is to be exhibited. It is not a report on investigations of the languages spoken by Englishmen or Frenchmen in the mass; no lexicographer ever reported such a division. One consequence of the Dichotomy is that every member of the class of individuals truly describable by any given proper predicate has something in common with every other member.

§2. In the *Tractatus* Wittgenstein shows no sign of dissatisfaction with the Traditional Dichotomy. But in the *Blue Book* he begins to express discontent with this feature of the model. The tendency to look for something common to all entities we commonly subsume under a general term is held to be one of several primitive, too simple ideas of the structure of language.² His rejection of the Traditional Dichotomy as inadequate to model English is given its most explicit form in *Philosophical Investigations*, §§66–67:

66. Consider for example the proceedings that we call 'games'. . . . What is common to them all? Don't say: 'There *must* be something in common or they would not be called "games" '—but *look and see* whether there is anything common to all. For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. . . . And the result of this examination is: we see a complicated network of similarities overlapping and cross-crossing: sometimes overall similarities, sometimes similarities of detail.

67. I can think of no better expression to characterize these similarities than 'family resemblances'. . . . And I shall say: 'games' form a family.³

§3. It is clear enough that Wittgenstein is rejecting the Traditional Dichotomy by way of denying the consequence mentioned in §1 above. This denial can be expressed in the doctrine:

There can be perfectly proper predicates—indeed "game" is one—which apply to objects having no one thing common to them all.

But the doctrine remains opaque so long as the criterion for *possession of something in common* is left unspecified.

The old test:

a and *b* have something in common just in case they are both truly describable by some one proper predicate,

is part and parcel of the position Wittgenstein is repudiating. A new criterion must therefore be furnished. Unhappily, Wittgenstein nowhere provides one, and as a result, his claim that predicates may apply in virtue of family resemblances among objects is insufficiently definite.

§4. The material specification of the test for what two given prime ministers have in common will differ from that for what is shared by two chess boards. Nevertheless, in the relation between the

¹ Explicit forms of this doctrine are to be found in Hobbes, *Elements of Philosophy*, ed. Molesworth (London, 1839), vol. I, pp. 23ff, and in J. S. Mill, *A System of Logic*, 9 ed. (London, 1875), vol. I, pp. 155ff., for example. In implicit forms, it is endemic.

² Wittgenstein, *Preliminary Studies . . . Generally Known as The Blue and Brown Books* (Oxford, 1960), p. 17.

³ Wittgenstein, *Philosophical Investigations* (Oxford, 1953), p. 32.

pair of prime ministers and what they share on the one hand, and the pair of chess boards and what *they* share on the other, there must lie some reason for claiming that both pairs are pairs of objects having something in common—a formal specification of the test must be found which will cover both cases of possessing something in common.⁴

One suggestion that springs naturally to mind is:

a and *b* have something in common (viz. *F*-hood), just in case they are indistinguishable with respect to *F*.

This suggestion is to be rejected. It is altogether too stringent. It requires that we deny of two postage stamps that they have their color in common if we can distinguish between them with respect to color. It requires that we deny of two so-called 4-inch nails that they have a common length not only if we find that one is 3.98 inches long and the other 4.02 inches, but also if we somehow discover that they differ in length by some amount which falls within the limits of quantitative discrimination, so that we know they differ but cannot say by how much.

The objection to this stringent criterion is that it results in the something in common/nothing in common distinction coinciding with another. Under this criterion *a* and *b* have something in common if and only if they are jointly describable by an exact predicate (one which does not admit of satisfaction by a range of cases); they have nothing in common just in case they are jointly describable by an inexact predicate (one which does). If there is no other predicate to hand, "being very like either *a* or *b*" is always available as an inexact predicate for this latter case.

The proposed criterion thus renders otiose the something/nothing in common distinction. Worse, it reduces Wittgenstein's claim that some predicates apply to objects having nothing in common to the claim that some predicates are inexact, a much less radical position. Worse still, it makes Wittgenstein contradict himself, for the very "overlapping and criss-crossing similarities" in virtue of which the family resemblance predicates are said to apply are themselves as often as not similarities picked out by inexact predicates. But Wittgenstein is clearly contrasting the whole family, all members

of which have no one thing in common, with pairs of its members, which do have things in common—things which may be picked out by inexact predicates. "Competitive," or "recreational," or "played with a ball" are examples for the family "game."

§5. The criterion for *a* and *b*'s possession of something in common just considered was too exclusive. We have already noticed, in §3, that

(a) jointly describable by some one proper predicate,

is too inclusive a criterion, for it would include even the families to the members of which Wittgenstein explicitly denies any single common possession.

Nevertheless, (a) does represent a necessary condition for the possession of something in common. For if *a* and *b* do have something in common, and there is no predicate at hand whereby they are jointly describable, we can introduce a predicate *F* such that *Fx* if and only if *x* has whatever it is that *a* and *b* share. *F* will then jointly describe *a* and *b*.

Some further restriction must therefore be applied to (a) to obtain a satisfactory necessary and sufficient condition for possession of something in common.

Wittgenstein's suggestion, that a predicate may apply to a class all members of which have no one thing in common, is not to be reduced to a trivial falsehood. Accordingly, the restriction to be placed on (a) must be powerful enough to exclude as expressing something in common predicates of the following classes:

- (i) Presence-functionally complex predicates. For all games are either games or whales, and all games are either indoor games or outdoor games.
- (ii) Negative predicates. For all games are non-prime-ministers.
- (iii) Non-extensible predicates. By this I mean predicates whose applicability cannot be extended to a newly encountered candidate on the basis of examination of that candidate alone, but essentially requires examination of something else as well. Examples describing all games would be "called 'game' by accurate speakers of English," "mentioned in the Complete World Encyclopedia of Games (if such there be)," or

⁴ Perhaps "having something in common" is a family resemblance two-place predicate, so that cases of having something in common have nothing in common. Then the required formal criterion will have a complex specification in terms of possession of something in common. I hold that the criterion developed in §§5, 6 below shows that this situation does not obtain. Rather, we need not so consider "having something in common"; and as generality is a virtue in logic, and all generalizations concerning families admit of exceptions, this is just as well.

"of less importance to heavy industry than cheap power."

- (iv) *Ipsa facto* predicates, i.e., predicates which apply where and insofar as the family resemblance predicates in question apply. All games, for example, fall in the complement of the complement of the class of games.

§6. The task of finding which restriction on criterion (a) will yield a satisfactory sufficient condition for possession of something in common seems now to be best tackled by seeking a rationale for the disqualification of the above four classes of predicate as predicates whose joint applicability to *a* and *b* is enough for *a* and *b* to have something in common.

To urge any one of the predicates from these four classes as a counter-example refuting Wittgenstein's claim that games have no one thing in common has an air of frivolity. Our rejection of such putative counter-examples does not seem a mere *ad hoc* rejection motivated solely by a desire to see some of Wittgenstein's remarks turn out to be true; on what grounds, then, do we disqualify predicates from these classes?

One good and sufficient reason is this, that the predicates in question can play no effective part in any explanation of how a predicate can properly be used to describe the many objects which exemplify it. The presence-functionally complex and the *ipso facto* predicates presuppose the success of this explanation, and the non-extensible and negative predicates are irrelevant.

To offer this reason involves appeal to the doctrines that the multiple applicability of predicates is something requiring justification, and that for some predicates this justification can be furnished by way of other linguistic entities, typically other predicates, whose propriety is already established. These doctrines lie behind the Traditional Dichotomy, in which all definable terms can be justified by reference to the constituents of the definiens. Explicit definition by a set of severally necessary and jointly sufficient conditions provides a clear case of the intra-linguistic justifications under discussion.

The doctrines in turn involve the notion of an epistemic hierarchy among predicates, such that those above can be justified by appeal to those below. Those predicates for which no intra-linguistic justification is possible (predicates of degree 0) are deemed to be *basic* predicates. For reasons furnished by Wittgenstein himself, any such epistemic order must be considered to be dependent upon, and

hence relative to, the total linguistic context in which it is found, and in particular to be relative to the stock of predicates available, and the capacities, choices, and purposes of the language-using group.

When cast in their most explicit form, all intra-linguistic justifications of predicates will appeal solely to predicates basic for the context in which the justifications are offered. Thus we might establish the propriety of the predicate "ice hockey" by appeal to the already established predicates "form of hockey," "played," and "on ice"; these latter having been established by appeal to such predicates as "curved stick," "hard ball," and "grassed field"; and these in turn established by appeal to the propriety of their conjunctive members.

Let us suppose that the possibility and necessity of such an hierarchical order among predicates is accepted.

The proposed restriction on (a) is then this:

- (b) *a* and *b* have something in common just in case they are both describable by some one basic predicate.

Expression of something in common is thus tied to furthering the explanation of the applicability of a predicate.

This suggestion not only supplies a single rationale for the disqualification of the complex, negative, non-extensible, and *ipso facto* predicates; it further neatly accounts for the contrast made between "game" and the predicates expressing the marks of game-hood, with respect to picking out something in common. For the latter can and do play a part in explaining the applicability of "game."

Even on the strengthened condition (b), all positive predicates recognized in the Traditional Dichotomy count as expressing possession of something in common, so we have not in this explication robbed Wittgenstein of a target for his denials.

A pleasant consequence of the relativity in the notion of basic predicate is that it makes *a*'s and *b*'s having something in common relative to the linguistic context in which the question arises, hence relative to the questioner's purposes in asking. In this, the proposed criterion accurately reproduces one feature of our use of the expression "something in common."

§7. Let the actual and possible individuals to which a predicate *F* applies be dubbed *F*'s *reference class*.

Wittgenstein's doctrine of family resemblance predicates can now be formulated as follows:

There are some proper predicates which have a reference class such that:

- (1) There is no one basic predicate which applies to every member.
- (2) Basic predicates do, however, apply to every member of various "overlapping and criss-crossing" subclasses of the reference class.
- (3) The predicate applies to the whole reference class in virtue of the applicability of these basic predicates to its subclasses.

The peculiarity of family resemblance predicates, on this account, is not merely that they admit of a range of cases in their reference class, but that further, we have the means to describe the relations between members falling within this range, in virtue of which the predicate is applicable to them all. To claim this is not to claim that in the case of family resemblance predicates we decide that they apply by first assuring ourselves of the applicability of a suitable set of basic predicates—we identify cases of "game" or "platypus" or "romanesque church" as directly and immediately as we identify cases of "curve," "magenta," or "strident." The claim is rather that for instances of the former trio there are linguistic resources available to us in meeting any challenge to show that a putative case is indeed a genuine case. In this, family resemblance predicates differ from basic ones. They differ from defined terms in that the challenge to show a case genuine is met in ways that vary systematically from instance to instance.

The Wittgensteinian doctrine as reformulated reproduces an error or omission in his remarks. All games are events: they all take place within a specifiable region of space, and they all last some time. They are all activities; they all have a beginning. Games all have at least one participant. These are all features which could be appealed to in explaining the applicability of "game," and where this is possible, they all count as things games have in common.

On their own, however, these predicates could not furnish a sufficient explanation of the applicability of "game"; even taken together, they would not suffice. In the claim that games have nothing in common, *nothing in common* is to be read as *nothing in common sufficient for the predicate's use*. And (1) above is to be amended to read:

- (1a) There is no set of basic predicates which applies

to every member, and in virtue of which alone the predicate applies.

§8. Until (2) above is made much more explicit, it will not be possible definitely to identify any predicate as a family resemblance predicate; nor will the information that a given predicate applies in virtue of family resemblances among its members be specific enough to be worth having. The necessary explication can be achieved by forming a description of the reference class and placing conditions on the form it must take.

We consider only individuals which are clear instances of the predicate under review—others are accommodated by an account of acceptable deviation from the norm.

The *reference class description* for F is made by describing members of a representative sample of F 's indefinitely large reference class. As the claim that any given sample is indeed representative can have no formal certitude, the claim that F is a family resemblance predicate (or indeed a predicate of any other species) can never be more than provisional.

The members of the chosen representative sample are arbitrarily ordered, and the *characteristics* (properties expressed by basic predicates) in virtue of which each member belongs to the reference class of F are listed in bracketed sets. The list for each individual is terminated and enclosed in brackets when it lists characteristics sufficient for F -hood. Thus:

Representative Sample	Reference Class Description
a	$(N_1, N_2 \dots N_n, M_1, M_2 \dots M_m)$
b	$(N_1 \dots N_n, M_m' \dots M_n)$
c	$(N_1 \dots M_p)$
d	$(N_1 \dots M_q)$

Where the bracketed set describing some individual k reproduces one for an individual earlier in the ordering of the representative sample, it is deleted. The characteristics necessary for F -hood, here labeled $N_1 \dots N_n$, occur in all bracketed sets. These are extracted and placed at the head of the reference class description:

$$(N_1, N_2, \dots, N_n) \{ (M_1 \dots M_m) (M_m' \dots M_n) \\ (M_n' \dots M_p) (M_p' \dots M_q) \dots \}$$

The characteristics labeled $M_1 \dots M_q \dots$ are relevant but not necessary characteristics of F -hood, as competitiveness is of games, or four-leggedness of cats. Such characteristics are com-

monly known as *marks* of *F*-hood. The necessary characteristics form the *N*-set of *F*'s reference class description, the remaining bracketed sets of marks form together its *M*-set.

Two complications must here be noticed. First, the reference class description must include characteristics, not only of members of the representative sample, but in addition, characteristics of other *relevant individuals* (individuals, not themselves being described, whose properties are relevant to the described individual's being an *F*). Thus "played with spherical balls" is a mark of snooker, but it is not basic, and so cannot occur in the reference class description of "snooker." "Spherical" is basic, but not a mark of snooker. The predicate "spherical," tagged to indicate that it describes the balls as relevant individuals, must be admitted to the reference class description of "snooker."

Second, this array requires that all characteristics be of the "all or nothing" kind, either completely present or completely absent from every individual. Predicates admitting degree (e.g., "warm") must therefore be eliminated in favor of a suitable number of quantity-indicating predicates (e.g., 5°C, . . . 25°C), each one of which is always simply present or absent. Neither of these complications involves any essential change.

§9. Now, *F*'s logical status is reflected in the form of its reference class description. For *F* to be a family resemblance predicate, *F*'s reference class description must fulfill the following conditions:

- (A) The *M*-set must not be null. For otherwise, it follows from the rules given for constructing the reference class description that possession of the *N*-set characteristics alone is sufficient for *F*-hood, and that hence *F* is a basic predicate or one conjunctively definable in terms of basic predicates, according as there are one or more members in the *N*-set. But this conflicts directly with condition (1a) of §7 above.
- (B) The bracketed sets of the *M*-set must not be independent of one another. They must, that is, contain members common to many other bracketed sets. For otherwise, *F* is a disjunctive

predicate (e.g., fire engine-or-ice block), whose members form two clearly separate groups. Such predicates violate the overlapping and crisscrossing requirement (2) of §7.

- (C) Every mark must occur in at least two bracketed sets. This is a precise requirement for *overlapping*.
- (D) No two logically independent marks are present together in every bracketed set in which either occurs. This parallels (C) for *crisscrossing*.

Requirements (A)–(D) specify minimum necessary and jointly sufficient conditions for *F*'s being a family resemblance predicate. For any such predicates in actual use, we require that they make a significant classification, that they should in some way single out a class members of which merit a like description. The family must, so to speak, be adequately closely knit.

These additional requirements can be expressed as further conditions on the reference class description. Among other conditions,

All bracketed sets must be of a certain minimum size,

They must all be of approximately equal size,

and

Each mark must occur in a certain minimum number of bracketed sets.

However, these minima for a happy family admit of no unique specification. We quite properly vary them from situation to situation, and predicate to predicate. We could not, for example, give to each mark a number correlated with its importance as a mark of *F*-hood, then specify that *a* is to be counted an *F* if and only if the sum total of the numbers of its characteristics is no less than some specified figure. For the minimum target figure would have to be varied from situation to situation and predicate to predicate. Conversely, recognizing this variation in the degree of coherence demanded of the families of which we are prepared to speak, we might use such a numerical technique to derive a measure of required coherence in a given situation.⁵

§10. We are now in a position to note some of the

⁵ Professor G. E. Hughes of Victoria University of Wellington has protested that any attempt to provide precise criteria for *something in common*, and any attempt to state express conditions for the applicability of "family resemblance predicate" does violence to Wittgenstein's intention, which so far from being that of proposing another label for predicates and hence another excuse for failing to examine the relations holding between members of the reference class, was that of encouraging us to a detailed examination, without preconceived notions, of each particular case. To this it must be replied that if we attempt the examination without a classifying model to bring to bear, we will fail to see the wood for the trees. And this being so, it is better to use a subtle model than a crude one. And furthermore, these very comments of Wittgenstein's provide the foundation for a most necessary elaboration of the model we employ. And that finally, the recognition of many abstract possibilities makes it less likely that we will be blind to the complexities of any concrete case.

salient features of family resemblance predicates.

A predicate is not a family resemblance predicate in any absolute sense, but only as relative to a given linguistic context embracing as variables the vocabulary, purposes, capacities, and choices of a language-using group. This follows from the relative nature of the basic status of the predicates used in the reference class description.

For any given linguistic context, not every predicate can be a family resemblance predicate. For the reference class description is made using basic predicates whose own reference class descriptions have *ex hypothesi* an empty *M*-set. The basic predicates, therefore, cannot be family resemblance terms. However, there seems to be no *a priori* ground for claiming that, for any particular predicate, there could be no linguistic context in which it was a family resemblance term. But of course in any such context there would have to be other predicates which were not family resemblance predicates.

This result stands in direct contradiction to the claim of Mr. Renford Bambrough in his paper "Universals and Family Resemblances."⁶ *F* is a family resemblance predicate, according to Bambrough, if *F*'s have nothing in common other than being *F*'s. This he claims to be true for every *F*. He disallows "Every brother has both maleness and sibling-hood in common with every other brother" as a refutation, on the ground that having either maleness or sibling-hood is not something *other* than having the property of being a brother. This is idiosyncratic enough. But the position is even worse with other cases:

Every Congressman is either a Senator or a Member of the House. According to Bambrough,

- (1) Congressmen have Congressmanhood in common and nothing else,
- (2) Senators have Senatorhood in common and nothing else.

Now, either Senators do not have Congressmanhood in common, which contradicts (1), or they *do* have Congressmanhood in common, which contradicts (2), or being a Congressman is not something other than being a Senator.

As this last is plainly false, it is reassuring that our result is in conflict with Bambrough's.

It follows that the notion of family resemblance cannot of itself solve the problem of universals for any given linguistic context. At best, by showing

that a further large class of predicates admit of justifying explanation by way of appeal to other predicates, we reduce the number of predicates requiring a separate extra-linguistic account.

§11. Family resemblance predicates are not eliminable in favor of any presence-functional combination of their marks. Consider a newly encountered candidate for membership in *F*'s reference class. Its description differs from any yet found in the representative sample, yet contains previously encountered marks in an overlapping and crisscrossing way, together with new ones. Where *F* is a family resemblance term, this new candidate will be counted a member of *F*'s reference class; if *F* is a disjunctive presence-function of a previously specified and finite list of marks, the new candidate must be excluded. Consequently, the one cannot replace the other, for the two are not even necessarily extensionally equivalent.

Furthermore, family resemblance predicates admit of a new class of borderline cases, namely individuals with a doubtfully sufficient set of relevant characteristics. But if *F* is a disjunction of combinations of characteristics, there can be no doubtful cases of this kind, for every such case must already have been decided in advance.

Still further, family resemblance predicates have an open texture in that it is not laid down in advance precisely which characteristics will come to be considered relevant, or which precedents will be followed concerning how tightly knit the family must be. But for presence-functional combinations of the marks, both these matters are so determined. These results parallel Professor Gasking's remarks on the relations between truths about clusters and truths about their members.⁷

§12. That family resemblance predicates are an intelligible possibility is clear. That some English terms conform closely to this model has been shown by Wittgenstein himself for many psychological terms. Many terms from natural history, and the discussion of human affairs (for example "fish," "baroque symphony," or "sovereign parliament") can likewise be exhibited as satisfying our conditions (A)-(D). As an adequate model for English, the Traditional Dichotomy is to be rejected. There is, indeed, no reason why there may not be many more differently patterned reference class descriptions to which proper predicates may correspond, and hence many species of predicate.

Nonetheless, family resemblance predicates are

⁶ *Proceedings of the Aristotelian Society*, vol. 61 (1960-61), pp. 207-222.

⁷ D. Gasking, "Clusters," *Australasian Journal of Philosophy*, vol. 38 (1960), pp. 1-36.

in some respects defective. Insofar as *F* and *G* admit of borderline cases, the truth-value of "All *F*'s are *G*'s" is indeterminate. In all cases where we wish to make generalizations, therefore, predicates more liable to borderline cases are to be replaced by others, making new classifications, which are less liable to embarrassing border disputes. But as already noted above, family resemblance predicates are heir to at least one more dimension of borderline case than either the basic, or the defined predicates countenanced by the Traditional Dichotomy.

Where *F* is a family resemblance predicate, the generalization "All *F*'s are *G*'s" admits of exceptions even where *G* is a criterial mark for *F*. With defined terms, this is not so. Accordingly, where we wish to make generalizations in the confidence that they admit of no exceptions, defined terms are to be preferred to family resemblance terms, other things being equal.

Most seriously, in making a classification by a family resemblance predicate, we are making a

division without reference to any natural ground it might have. For otherwise we would be classifying on that basis and all members of the reference class would have something in common, to wit, that which gives rise to the observed pattern of surface phenomena. For the discussions of human affairs, this is perhaps of no great moment. But it is our faith that all divisions in nature, save the most fundamental, have a ground. Natural-history categories, such as *sand dune*, or *tree*, or *fish*, play no essential role in developed sciences. In chemistry there has been a long search for adequate replacements for the family resemblance predicates picking out natural substances.

The Traditional Dichotomy not only offered a model, it also held up as an ideal a system of classification which fits the grounded divisions of nature. This in part explains the Dichotomy's long life. The ideal is not a paltry one; we should not rest content until family resemblance predicates, admittedly intelligible, have been banished from our sciences.

University of Melbourne

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston	James M. Edie	Richard H. Popkin
Alan R. Anderson	Peter Thomas Geach	Wesley C. Salmon
Kurt Baier	Adolf Grünbaum	George A. Schrader
Lewis W. Beck	Carl G. Hempel	Wilfrid Sellars
Richard B. Brandt	Jaakko Hintikka	J. J. C. Smart
Roderick M. Chisholm	Raymond Klibansky	Wolfgang Stegmüller
L. Jonathan Cohen	Benson Mates	Manley H. Thompson, Jr.
James Collins	John A. Passmore	G. H. von Wright
Michael Dummett	Günther Patzig	John W. Yolton

VOLUME 2/NUMBER 4

CONTENTS

OCTOBER 1965

I. J. L. MACKIE: <i>Causes and Conditions</i>	245	V. J. E. LLEWELYN: <i>Propositions as Answers</i>	305
II. W. C. SALMON, S. F. BARKER, H. E. KYBURG, JR.: <i>Symposium on Inductive Evidence</i>	265	VI. CLEMENT DORE: <i>Seeming to See</i>	312
III. C. S. CHIHARA AND J. A. FODOR: <i>Operationalism and Ordinary Language: A Critique of Wittgenstein</i>	281	VII. FRANCIS J. COLEMAN: <i>Can a Smell or a Taste or a Touch Be Beautiful?</i>	319
IV. BRIAN ELLIS: <i>A Vindication of Scientific Inductive Practices</i>	296	VIII. <i>Corrigenda</i>	325

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

EDITORIAL COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

REPRINTS

Each author will receive gratis two copies of the issue in which his article appears. Reprints can be purchased at low charges through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* for individual subscribers is six dollars and the price *per annum* for institutions is ten dollars. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues are sold at the rate of two dollars to individuals, and three dollars to institutions. Correspondence regarding subscription and back orders may be addressed directly to the publisher (Basil Blackwell, Oxford, England).

* * *

Through 1965, the *American Philosophical Quarterly* is published quarterly in January, April, July, and October by the University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, Pennsylvania, 15213.

Second-class postage paid at Pittsburgh, Pennsylvania.



I. CAUSES AND CONDITIONS

J. L. MACKIE

ASKED what a cause is, we may be tempted to say that it is an event which precedes the event of which it is the cause, and is both necessary and sufficient for the latter's occurrence; briefly, that a cause is a necessary and sufficient preceding condition. There are, however, many difficulties in this account. I shall try to show that what we often speak of as a cause is a condition not of this sort, but of a sort related to this. That is to say, this account needs modification, and can be modified, and when it is modified we can explain much more satisfactorily how we can arrive at much of what we ordinarily take to be causal knowledge; the claims implicit within our causal assertions can be related to the forms of the evidence on which we are often relying when we assert a causal connection.

§ 1. SINGULAR CAUSAL STATEMENTS

Suppose that a fire has broken out in a certain house, but has been extinguished before the house has been completely destroyed. Experts investigate the cause of the fire, and they conclude that it was caused by an electrical short-circuit at a certain place. What is the exact force of their statement that this short-circuit caused this fire? Clearly the experts are not saying that the short-circuit was a necessary condition for this house's catching fire at this time; they know perfectly well that a short-circuit somewhere else, or the overturning of a lighted oil stove, or any one of a number of other things might, if it had occurred, have set the house on fire. Equally, they are not saying that the short-circuit was a sufficient condition for this house's catching fire; for if the short-circuit had occurred, but there had been no inflammable material nearby, the fire would not have broken out, and even given both the short-circuit and the inflammable material, the fire would not have occurred if, say, there had been an efficient automatic sprinkler at just the right spot. Far from being a condition both necessary and sufficient for

the fire, the short-circuit was, and is known to the experts to have been, neither necessary nor sufficient for it. In what sense, then, is it said to have caused the fire?

At least part of the answer is that there is a set of conditions (of which some are positive and some are negative), including the presence of inflammable material, the absence of a suitably placed sprinkler, and no doubt quite a number of others, which combined with the short-circuit constituted a complex condition that was sufficient for the house's catching fire—sufficient, but not necessary, for the fire could have started in other ways. Also, of *this* complex condition, the short-circuit was an indispensable part: the other parts of this condition, conjoined with one another in the absence of the short-circuit, would not have produced the fire. The short-circuit which is said to have caused the fire is thus an indispensable part of a complex sufficient (but not necessary) condition of the fire. In this case, then, the so-called cause is, and is known to be, an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result. The experts are saying, in effect, that the short-circuit is a condition of this sort, that it occurred, that the other conditions which conjoined with it form a sufficient condition were also present, and that no other sufficient condition of the house's catching fire was present on this occasion. I suggest that when we speak of the cause of some particular event, it is often a condition of this sort that we have in mind. In view of the importance of conditions of this sort in our knowledge of and talk about causation, it will be convenient to have a short name for them: let us call such a condition (from the initial letters of the words italicized above), an *INUS* condition.¹

This account of the force of the experts' statement about the cause of the fire may be confirmed by reflecting on the way in which they will have reached this conclusion, and the way in which anyone who disagreed with it would have to

¹ This term was suggested by D. C. Stove who has also given me a great deal of help by criticizing earlier versions of this article.

challenge it. An important part of the investigation will have consisted in tracing the actual course of the fire; the experts will have ascertained that no other condition sufficient for a fire's breaking out and taking this course was present, but that the short-circuit did occur and that conditions were present which in conjunction with it were sufficient for the fire's breaking out and taking the course that it did. Provided that there is some necessary and sufficient condition of the fire—and this is an assumption that we commonly make in such contexts—anyone who wanted to deny the experts' conclusion would have to challenge one or another of these points.

We can give a more formal analysis of the statement that something is an INUS condition. Let ' A ' stand for the INUS condition—in our example, the occurrence of a short-circuit at that place—and let ' B ' and ' \bar{C} ' (that is, 'not- C ', or the absence of C) stand for the other conditions, positive and negative, which were needed along with A to form a sufficient condition of the fire—in our example, B might be the presence of inflammable material, \bar{C} the absence of a suitably placed sprinkler. Then the conjunction ' ABC ' represents a sufficient condition of the fire, and one that contains no redundant factors; that is, ABC is a minimal sufficient condition for the fire.² Similarly, let \overline{DEF} , \overline{GHI} , etc., be all the other minimal sufficient conditions of this result. Now provided that there is some necessary and sufficient condition for this result, the disjunction of all the minimal sufficient conditions for it constitutes a necessary and sufficient condition.³ That is, the formula " ABC or \overline{DEF} or \overline{GHI} or . . ." represents a necessary and sufficient condition for the fire, each of its disjuncts, such as ' ABC ', represents a minimal sufficient condition, and each conjunct in each minimal sufficient con-

dition, such as ' A ', represents an INUS condition. To simplify and generalize this, we can replace the conjunction of terms conjoined with ' A ' (here ' BC ') by the single term ' X ', and the formula representing the disjunction of all the other minimal sufficient conditions—here " \overline{DEF} or \overline{GHI} or . . ."—by the single term ' Y '. Then an INUS condition is defined as follows:

A is an INUS condition of a result P if and only if, for some X and for some Y , (AX or Y) is a necessary and sufficient condition of P , but A is not a sufficient condition of P and X is not a sufficient condition of P .

We can indicate this type of relation more briefly if we take the provisos for granted and replace the existentially quantified variables ' X ' and ' Y ' by dots. That is, we can say that A is an INUS condition of P when (A . . . or . . .) is a necessary and sufficient condition of P .

(To forestall possible misunderstandings, I would fill out this definition as follows.⁴ First, there could be a set of minimal sufficient conditions of P , but no necessary conditions, not even a complex one; in such a case, A might be what Marc-Wogau calls a moment in a minimal sufficient condition, but I shall not call it an INUS condition. I shall speak of an INUS condition only where the disjunction of all the minimal sufficient conditions is also a necessary condition. Secondly, the definition leaves it open that the INUS condition A might be a conjunct in each of the minimal sufficient conditions. If so, A would be itself a necessary condition of the result. I shall still call A an INUS condition in these circumstances: it is not part of the definition of an INUS condition that it should *not* be necessary, although in the standard cases, such as that

² The phrase "minimal sufficient condition" is borrowed from Konrad Marc-Wogau, "On Historical Explanation," *Theoria*, vol. 28 (1962), pp. 213–233. This article gives an analysis of singular causal statements, with special reference to their use by historians, which is substantially equivalent to the account I am suggesting. Many further references are made to this article, especially in n. 9 below.

³ Cf. n. 8 on p. 227 of Marc-Wogau's article, where it is pointed out that in order to infer that the disjunction of all the minimal sufficient conditions will be a necessary condition, "it is necessary to presuppose that an arbitrary event C , if it occurs, must have sufficient reason to occur." This presupposition is equivalent to the presupposition that there is some (possibly complex) condition that is both necessary and sufficient for C .

It is of some interest that some common turns of speech embody this presupposition. To say "Nothing but X will do," or "Either X or Y will do, but nothing else will," is a natural way of saying that X , or the disjunction (X or Y), is a necessary condition for whatever result we have in mind. But taken literally these remarks say only that there is no sufficient condition for this result other than X , or other than (X or Y). That is, we use to mean "a necessary condition" phrases whose literal meanings would be "the only sufficient condition," or "the disjunction of all sufficient conditions." Similarly, to say that Z is "all that's needed" is a natural way of saying that Z is a sufficient condition, but taken literally this remark says that Z is the only necessary condition. But, once again, that the only necessary condition will also be a sufficient one follows only if we presuppose that some condition is both necessary and sufficient.

⁴ I am indebted to the referees for the suggestion that these points should be clarified.

sketched above, it is not in fact necessary.⁵ Thirdly, the requirement that X by itself should not be sufficient for P insures that A is a nonredundant part of the sufficient condition AX ; but there is a sense in which it may not be strictly necessary or indispensable even as a part of *this* condition, for it may be replaceable: for example KX might be another minimal sufficient condition of P .⁶ Fourthly, it is part of the definition that the minimal sufficient condition, AX , of which A is a nonredundant part, is not also a necessary condition, that there is another sufficient condition Y (which may itself be a disjunction of sufficient conditions). Fifthly, and similarly, it is part of the definition that A is not by itself sufficient for P . The fourth and fifth of these points amount to this: I shall call A an INUS condition only if there are terms which actually occupy the places occupied by ' X ' and ' Y ' in the formula for the necessary and sufficient condition. However, there may be cases where there is only one minimal sufficient condition, say AX . Again, there may be cases where A is itself a minimal sufficient condition, the disjunction of all minimal sufficient conditions being $(A \text{ or } Y)$; again, there may be cases where A itself is the only minimal sufficient condition, and is itself both necessary and sufficient for P . In any of these cases, as well as in cases where A is an INUS condition, I shall say that A is *at least an INUS condition*. As we shall see, we often have evidence which supports the conclusion that something is *at least an INUS condition*; we may or may not have other evidence which shows that it is *no more than an INUS condition*.)

I suggest that a statement which asserts a singular causal sequence, of such a form as " A caused P ," often makes, implicitly, the following claims:

- (i) A is at least an INUS condition of P —that is,

there is a necessary and sufficient condition of P which has one of these forms: $(AX \text{ or } Y)$, $(A \text{ or } Y)$, AX , A .

(ii) A was present on the occasion in question.

(iii) The factors represented by the ' X ', if any, in the formula for the necessary and sufficient condition were present on the occasion in question.

(iv) Every disjunct in ' Y ' which does not contain ' A ' as a conjunct was absent on the occasion in question. (As a rule, this means that whatever ' Y ' represents was absent on this occasion. If ' Y ' represents a single conjunction of factors, then it was absent if at least one of its conjuncts was absent; if it represents a disjunction, then it was absent if each of its disjuncts was absent. But we do not wish to exclude the possibility that ' Y ' should be, or contain as a disjunct, a conjunction one of whose conjuncts is A , or to require that *this* conjunction should have been absent.⁷)

I do not suggest that this is the whole of what is meant by " A caused P " on any occasion, or even that it is a part of what is meant on every occasion: some additional and alternative parts of the meaning of such statements are indicated below.⁸ But I am suggesting that this is an important part of the concept of causation; the proof of this suggestion would be that in many cases the falsifying of any one of the above-mentioned claims would rebut the assertion that A caused P .

This account is in fairly close agreement, in substance if not in terminology, with at least two accounts recently offered of the cause of a single event.

Konrad Marc-Wogau sums up his account thus:

when historians in singular causal statements speak of a cause or the cause of a certain individual event β , then what they are referring to is another individual event α which is a moment in a minimal sufficient and at the same time necessary condition *post factum* β .⁹

⁵ Special cases where an INUS condition is also a necessary one are mentioned at the end of § 3.

⁶ This point, and the term "nonredundant," are taken from Michael Scriven's review of Nagel's *The Structure of Science*, in *Review of Metaphysics*, 1964. See especially the passage on p. 408 quoted below.

⁷ See example of the wicket-keeper discussed below.

⁸ See §§ 7, 8.

⁹ See pp. 226–227 of the article referred to in n. 2 above. Marc-Wogau's full formulation is as follows:

'Let 'msc' stand for minimal sufficient condition and 'nc' for necessary condition. Then suppose we have a class K of individual events a_1, a_2, \dots, a_n . (It seems reasonable to assume that K is finite; however even if K were infinite the reasoning below would not be affected.) My analysis of the singular causal statement: α is the cause of β , where α and β stand for individual events, can be summarily expressed in the following statements:

(1) $(EK) (K = \{a_1, a_2, \dots, a_n\})$;	(4) $(x) ((x \in K \wedge x \neq a_1) \supset x \text{ is not fulfilled when } \alpha \text{ occurs})$;
(2) $(x) (x \in K \equiv x \text{ msc } \beta)$;	(5) $\alpha \text{ is a moment in } a_1$.
(3) $(a_1 \vee a_2 \vee \dots \vee a_n) \text{ nc } \beta$;	

(3) and (4) say that a_1 is a necessary condition *post factum* for β . If a_1 is a necessary condition *post factum* for β , then every moment in a_1 is a necessary condition *post factum* for β , and therefore also α . As has been mentioned before (note 6) there is assumed to be a temporal sequence between α and β ; β is not itself an element in K ."

He explained his phrase "necessary condition *post factum*" by saying that he will call an event a_1 a necessary condition *post factum* for x if the disjunction " a_1 or a_2 or a_3 . . . or a_n " represents a necessary condition for x , and of these disjuncts only a_1 was present on the particular occasion when x occurred.

Similarly Michael Scriven has said:

Causes are *not* necessary, even contingently so, they are not sufficient—but they are, to talk that language, *contingently sufficient*. . . . They are part of a set of conditions that does guarantee the outcome, and they are non-redundant in that the rest of *this* set (which does not include all the other conditions present) is not alone sufficient for the outcome. It is not even true that they are relatively necessary, i.e., necessary with regard to that set of conditions rather than the total circumstances of their occurrence, for there may be several possible replacements for them which happen not to be present. There remains a ghost of necessity; a cause is a factor from a set of possible factors the presence of one of which (*any* one) is necessary in order that a set of conditions actually present be sufficient for the effect.¹⁰

There are only slight differences between these two accounts, or between each of them and that offered above. Scriven seems to speak too strongly when he says that causes are not necessary: it is, indeed, not part of the definition of a cause of this sort that it should be necessary, but, as noted above, a cause, or an INUS condition, may be necessary, either because there is only one minimal sufficient condition or because the cause is a moment in each of the minimal sufficient conditions. On the other hand, Marc-Wogau's account of a minimal sufficient condition seems too strong. He says that a minimal sufficient condition contains "only those moments relevant to the effect" and that a moment is relevant to an effect if "it is a necessary condition for β : β would not have occurred if this moment had not been present." This is less accurate than Scriven's statement that the cause only needs to be nonredundant.¹¹ Also, Marc-Wogau's requirement, in his account of a

necessary condition *post factum*, that only one minimal sufficient condition (the one containing a) should be present on the particular occasion, seems a little too strong. If two or more minimal sufficient conditions (say a_1 and a_2) were present, but a was a moment in each of them, then though neither a_1 nor a_2 was necessary *post factum*, a would be so. I shall use this phrase "necessary *post factum*" to include cases of this sort: that is, a is a necessary condition *post factum* if it is a moment in every minimal sufficient condition that was present. For example, in a cricket team the wicket-keeper is also a good batsman. He is injured during a match, and does not bat in the second innings, and the substitute wicket-keeper drops a vital catch that the original wicket-keeper would have taken. The team loses the match, but it would have won if the wicket-keeper had *both* batted *and* taken that catch. His injury was a moment in two minimal sufficient conditions for the loss of the match; either his not batting, or the catch's not being taken, would on its own have insured the loss of the match. But we can certainly say that his injury caused the loss of the match, and that it was a necessary condition *post factum*.

This account may be summed up, briefly and approximately, by saying that the statement " A caused P " often claims that A was necessary and sufficient for P in the circumstances. This description applies in the standard cases, but we have already noted that a cause is nonredundant rather than necessary even in the circumstances, and we shall see that there are special cases in which it may be neither necessary nor nonredundant.

§ 2. DIFFICULTIES AND REFINEMENTS¹²

Both Scriven and Marc-Wogau are concerned not only with this basic account, but with certain difficulties and with the refinements and complications that are needed to overcome them. Before dealing with these I shall introduce, as a refinement of my own account, the notion of a causal field.¹³

¹⁰ *Op. cit.*, p. 408.

¹¹ However, in n. 7 on pp. 222–233, Marc-Wogau draws attention to the difficulty of giving an accurate definition of "a moment in a sufficient condition." Further complications are involved in the account given in § 5 below of "clusters" of factors and the progressive localization of a cause. A condition which is minimally sufficient in relation to one degree of analysis of factors may not be so in relation to another degree of analysis.

¹² This section is something of an aside: the main argument is resumed in § 3.

¹³ This notion of a causal field was introduced by John Anderson. He used it, e.g., in "The Problem of Causality," first published in the *Australasian Journal of Psychology and Philosophy*, vol. 16 (1938), and reprinted in *Studies in Empirical Philosophy* (Sydney, 1962), pp. 126–136, to overcome certain difficulties and paradoxes in Mill's account of causation. I have also used this notion to deal with problems of legal and moral responsibility, in "Responsibility and Language," *Australasian Journal of Philosophy*, vol. 33 (1955), pp. 143–159.

This notion is most easily explained if we leave, for a time, singular causal statements and consider general ones. The question "What causes influenza?" is incomplete and partially indeterminate. It may mean "What causes influenza in human beings in general?" If so, the (full) cause that is being sought is a difference that will mark off cases in which human beings contract influenza from cases in which they do not; the causal field is then the region that is to be thus divided, *human beings in general*. But the question may mean, "Given that influenza viruses are present, what makes some people contract the disease whereas others do not?" Here the causal field is *human beings in conditions where influenza viruses are present*. In all such cases, the cause is required to differentiate, within a wider region in which the effect sometimes occurs and sometimes does not, the sub-region in which it occurs: this wider region is the causal field. This notion can now be applied to singular causal questions and statements. "What caused this man's skin cancer?"¹⁴ may mean "Why did this man develop skin cancer now when he did not develop it before?" Here the causal field is the career of this man: it is within this that we are seeking a difference between the time when skin cancer developed and times when it did not. But the same question may mean "Why did this man develop skin cancer, whereas other men who were also exposed to radiation did not?" Here the causal field is the class of men thus exposed to radiation. And what is the cause in relation to one field may not be the cause in relation to another. Exposure to a certain dose of radiation may be the cause in relation to the former field: it cannot be the cause in relation to the latter field since it is part of the description of that field, and being present throughout that field it cannot differentiate one sub-region of it from another. In relation to the latter field, the cause may be, in Scriven's terms, "Some as-yet-unidentified constitutional factor."

In our first example of the house which caught fire, the history of this house is the field in relation to which the experts were looking for the cause of the fire: their question was "Why did this house catch fire on this occasion, and not on others?" However, there may still be some indeterminacy in this choice of a causal field. Does this house,

considered as the causal field, include all its features, or all its relatively permanent features, or only some of these? If we take all its features, or even all of its relatively permanent ones, as constituting the field, then some of the things that we have treated as conditions—for example the presence of inflammable material near the place where the short-circuit occurred—would have to be regarded as parts of the field, and we could not then take them also as conditions which in relation to this field, as additions to it or intrusions into it, are necessary or sufficient for something else. We must therefore take the house, in so far as it constitutes the causal field, as determined only in a fairly general way, by only some of its relatively permanent features, and we shall then be free to treat its other features as conditions which do not constitute the field, and are not parts of it, but which may occur within it or be added to it. It is in general an arbitrary matter whether a particular feature is regarded as a condition (that is, as a possible causal factor) or as part of the field, but it cannot be treated in both ways at once. If we are to say that something happened to this house because of, or partly because of, a certain feature, we are implying that it would still have been *this* house, the house in relation to which we are seeking the cause of this happening, even if it had not had this particular feature.

I now propose to modify the account given above of the claims often made by singular causal statements. A statement of such a form as "*A* caused *P*" is usually elliptical, and is to be expanded into "*A* caused *P* in relation to the field *F*." And then in place of the claim stated in (i) above, we require this:

(ia) *A* is at least an *INUS* condition of *P* in the field *F*—that is, there is a condition which, given the presence of whatever features characterize *F* throughout, is necessary and sufficient for *P*, and which is of one of these forms: (*AX* or *Y*), (*A* or *Y*), *AX*, *A*.

In analyzing our ordinary causal statements, we must admit that the field is often taken for granted or only roughly indicated, rather than specified precisely. Nevertheless, the field in relation to which we are looking for a cause of this effect, or saying that such-and-such is a cause, may be definite enough for us to be able to say

¹⁴ These examples are borrowed from Scriven, *op. cit.*, pp. 409–410. Scriven discusses them with reference to what he calls a "contrast class," the class of cases where the effect did not occur with which the case where it did occur is being contrasted. What I call the causal field is the logical sum of the case (or cases) in which the effect is being said to be caused with what Scriven calls the contrast class.

that certain facts or possibilities are irrelevant to the particular causal problem under consideration, because they would constitute a shift from the intended field to a different one. Thus if we are looking for the cause, or causes, of influenza, meaning its cause(s) in relation to the field *human beings*, we may dismiss, as not directly relevant, evidence which shows that some proposed cause fails to produce influenza in rats. If we are looking for the cause of the fire in *this house*, we may similarly dismiss as irrelevant the fact that a proposed cause would not have produced a fire if the house had been radically different, or had been set in a radically different environment.

This modification enables us to deal with the well-known difficulty that it is impossible, without including in the cause the whole environment, the whole prior state of the universe (and so excluding any likelihood of repetition), to find a genuinely sufficient condition, one which is "by itself, adequate to secure the effect."¹⁵ It may be hard to find even a complex condition which was absolutely sufficient for this fire because we should have to include, as one of the negative conjuncts, such an item as the earth's not being destroyed by a nuclear explosion just after the occurrence of the suggested INUS condition; but it is easy and reasonable to say simply that such an explosion would, in more senses than one, take us outside the field in which we are considering this effect. That is to say, it may be not so difficult to find a condition which is sufficient in relation to the intended field. No doubt this means that causal statements may be vague, in so far as the specification of the field is vague, but this is not a serious obstacle to establishing or using them, either in science or in everyday contexts.¹⁶

It is a vital feature of the account I am suggesting that we can say that *A* caused *P*, in the sense

described, without being able to specify exactly the terms represented by '*X*' and '*Y*' in our formula. In saying that *A* is at least an INUS condition for *P* in *F*, one is *not* saying what other factors, along with *A*, were both present and nonredundant, and one is *not* saying what other minimal sufficient conditions there may be for *P* in *F*. One is not even claiming to be able to say what they are. This is in no way a difficulty: it is a readily recognizable fact about our ordinary causal statements, and one which this account explicitly and correctly reflects.¹⁷ It will be shown (in § 5 below) that this elliptical or indeterminate character of our causal statements is closely connected with some of our characteristic ways of discovering and confirming causal relationships: it is precisely for statements that are thus "gappy" or indeterminate that we can obtain fairly direct evidence from quite modest ranges of observation. On this analysis, causal statements implicitly contain existential quantifications; one can assert an existentially quantified statement without asserting any instantiation of it, and one can also have good reason for asserting an existentially quantified statement without having the information needed to support any precise instantiation of it. I can know that there is someone at the door even if the question "Who is he?" would floor me.

Marc-Wogau is concerned especially with cases where "there are two events, each of which independently of the other is a sufficient condition for another event." There are, that is to say, two minimal sufficient conditions, both of which actually occurred. For example, lightning strikes a barn in which straw is stored, and a tramp throws a burning cigarette butt into the straw at the same place and at the same time. Likewise for an historical event there may be more than one "cause," and each of them may, on its own, be

¹⁵ Cf. Bertrand Russell, "On the Notion of Cause," *Mysticism and Logic* (London, 1917), p. 187. Cf. also Scriven's first difficulty, *op. cit.*, p. 409: "First, there are virtually no known sufficient conditions, literally speaking, since human or accidental interference is almost inexhaustibly possible, and hard to exclude by specific qualification without tautology." The introduction of the causal field also automatically covers Scriven's third difficulty and third refinement, that of the contrast class and the relativity of causal statements to contexts.

¹⁶ J. R. Lucas, "Causation," *Analytical Philosophy*, ed. R. J. Butler (Oxford, 1962), pp. 57-59, resolves this kind of difficulty by an informal appeal to what amounts to this notion of a causal field: "... these circumstances [cosmic cataclysms, etc.] ... destroy the whole causal situation in which we had been looking for *Z* to appear ... predictions are not expected to come true when quite unforeseen emergencies arise."

¹⁷ This is related to Scriven's second difficulty, *op. cit.*, p. 409: "there still remains the problem of saying what the other factors are which, with the cause, make up the sufficient condition. If they can be stated, causal explanation is then simply a special case of subsumption under a law. If they cannot, the analysis is surely mythological." Scriven correctly replies that "a combination of the thesis of macro-determinism ... and observation-plus-theory frequently gives us the very best of reasons for saying that a certain factor combines with an unknown sub-set of the conditions present into a sufficient condition for a particular effect." He gives a statistical example of such evidence, but the whole of my account of typical sorts of evidence for causal relationships in §§ 5 and 7 below is an expanded defence of a reply of this sort.

sufficient.¹⁸ Similarly Scriven considers a case where

... conditions (perhaps unusual excitement plus constitutional inadequacies) [are] present at 4.0 P.M. that guarantee a stroke at 4.55 P.M. and consequent death at 5.0 P.M.; but an entirely unrelated heart attack at 4.50 P.M. is still correctly called the cause of death, which, as it happens, does occur at 5.0 P.M.¹⁹

Before we try to resolve these difficulties let us consider another of Marc-Wogau's problems: Smith and Jones commit a crime, but if they had not done so the head of the criminal organization would have sent other members to perform it in their stead, and so it would have been committed anyway.²⁰ Now in this case, if '*A*' stands for the actions of Smith and Jones, what we have is that *AX* is one minimal sufficient condition of the result (the crime), but $\overline{A}Z$ is another, and both *X* and *Z* are present. *A* combines with one set of the standing conditions to produce the result by one route: but the absence of *A* would have combined with another set of the standing conditions to produce the same result by another route. In this case we can say that *A* was a necessary condition *post factum*. This sample satisfies the requirements of Marc-Wogau's analysis, and of mine, of the statement that *A* caused this result; and this agrees with what we would ordinarily say in such a case. (We might indeed add that there was *also* a deeper cause—the existence of the criminal organization, perhaps—but this does not matter: our formal analyses do not insure that a particular result will have a unique cause, nor does our ordinary causal talk require this.) It is true that in this case we cannot say what will usually serve as an informal substitute for the formal account, that the cause, here *A*, was necessary (as well as sufficient) in the circumstances; for \overline{A} would have done just as well. We cannot even say that *A* was nonredundant. But this shows merely that a formal analysis may be superior to its less formal counterparts.

Now in Scriven's example, we might take it that the heart attack prevented the stroke from occurring. If so, then the heart attack is a necessary condition *post factum*: it is a moment in the only minimal sufficient condition that was present in full, for the heart attack itself removed some factor

that was a necessary part of the minimal sufficient condition which has the excitement as one of its moments. This is strictly parallel to the Smith and Jones case. Again it is odd to say that the heart attack was in any way necessary, since the absence of the heart attack would have done just as well: this absence would have been a moment in that other minimal sufficient condition, one of whose other moments was the excitement. Nevertheless, the heart attack was necessary *post factum*, and the excitement was not. Scriven draws the distinction, quite correctly, in terms of continuity and discontinuity of causal chains: "the heart attack was, and the excitement was not the cause of death because the 'causal chain' between the latter and death was interrupted, while the former's 'went to completion'." But it is worth noting that a break in the causal chain corresponds to a failure to satisfy the logical requirements of a moment in a minimal sufficient condition that is also necessary *post factum*.

Alternatively, if the heart attack did not prevent the stroke, then we have a case parallel to that of the straw in the barn, or of the man who is shot by a firing squad, and two bullets go through his heart simultaneously. In such cases the requirements of my analysis, or of Marc-Wogau's, or of Scriven's, are not met: each proposed cause is redundant and not even necessary *post factum*, though the disjunction of them is necessary *post factum* and nonredundant. But this agrees very well with the fact that we *would* ordinarily hesitate to say, of either bullet, that it caused the man's death, or of either the lightning or the cigarette butt that it caused the fire, or of either the excitement or the heart attack that it was the cause of death. As Marc-Wogau says, "in such a situation as this we are unsure also how to use the word 'cause'." Our ordinary concept of cause does not deal clearly with cases of this sort, and we are free to decide whether or not to add to our ordinary use, and to the various more or less formal descriptions of it, rules which allow us to say that where more than one at-least-INUS-condition, and its conjunct conditions, are present, each of them caused the result.²¹

The account thus far developed of singular causal statements has been expressed in terms of

¹⁸ *Op. cit.*, pp. 228–233.

¹⁹ *Op. cit.*, pp. 410–411: this is Scriven's fourth difficulty and refinement.

²⁰ *Op. cit.*, p. 232: the example is taken from P. Gardiner, *The Nature of Historical Explanation* (Oxford, 1952), p. 101.

²¹ Scriven's fifth difficulty and refinement are concerned with the direction of causation. This is considered briefly in § 8 below.

statements about necessity and sufficiency; it is therefore incomplete until we have added an account of necessity and sufficiency themselves. This question is considered in § 4 below. But the present account is independent of any particular analysis of necessity and sufficiency. Whatever analysis of these we finally adopt, we shall use it to complete the account of what it is to be an INUS condition, or to be at least an INUS condition. But in whatever way this account is completed, we can retain the general principle that at least part of what is often done by a singular causal statement is to pick out, as the cause, something that is claimed to be at least an INUS condition.

§ 3. GENERAL CAUSAL STATEMENTS

Many general causal statements are to be understood in a corresponding way. Suppose, for example, that an economist says that the restriction of credit causes (or produces) unemployment. Again, he will no doubt be speaking with reference to some causal field; this is now not an individual object, but a class, presumably economies of a certain general kind; perhaps their specification will include the feature that each economy of the kind in question contains a large private enterprise sector with free wage-earning employees. The result, unemployment, is something which sometimes occurs and sometimes does not occur within this field, and the same is true of the alleged cause, the restriction of credit. But the economist is not saying that (even in relation to this field) credit restriction is either necessary or sufficient for unemployment, let alone both necessary and sufficient. There may well be other circumstances which must be present along with credit restriction, in an economy of the kind referred to, if unemployment is to result; these other circumstances will no doubt include various negative ones, the absence of various counteracting causal factors which, if they were present, would prevent this result. Also, the economist will probably be quite prepared to admit that in an economy of this kind unemployment could be brought about by other combinations of circumstances in which the restriction of credit plays no part. So once again the claim that he is making is merely that the restriction of credit is, in economies of this kind, a nonredundant part of one sufficient condition for unemployment: that is, an INUS condition. The economist is probably assuming that there is some condition, no doubt a complex one, which is both necessary and

sufficient for unemployment in this field. This being assumed, what he is asserting is that, for some X and for some Y , (AX or Y) is a necessary and sufficient condition for P in F , but neither A nor X is sufficient on its own, where ' A ' stands for the restriction of credit, ' P ' for unemployment, and ' F ' for the field, economies of such-and-such a sort. In a developed economic theory the field F may be specified quite exactly, and so may the relevant combinations of factors represented here by ' X ' and ' Y '. (Indeed, the theory may go beyond statements in terms of necessity and sufficiency to ones of functional dependence, but this is a complication which I am leaving aside for the present.) In a preliminary or popular statement, on the other hand, the combinations of factors may either be only roughly indicated or be left quite undetermined. At one extreme we have the statement that (AX or Y) is a necessary and sufficient condition, where ' X ' and ' Y ' are given definite meanings; at the other extreme we have the merely existentially quantified statement that this holds for *some* pair X and Y . Our knowledge in such cases ordinarily falls somewhere between these two extremes. We can use the same convention as before, deliberately allowing it to be ambiguous between these different interpretations, and say that in any of these cases; where A is an INUS condition of P in F , ($A \dots$ or \dots) is a necessary and sufficient condition of P in F .

A great deal of our ordinary causal knowledge is of this form. We know that the eating of sweets causes dental decay. Here the field is human beings who have some of their own teeth. We do not know, indeed it is not true, that the eating of sweets by any such person is a sufficient condition for dental decay: some people have peculiarly resistant teeth, and there are probably measures which, if taken along with the eating of sweets, would protect the eater's teeth from decay. All we know is that sweet-eating combined with a set of positive and negative factors which we can specify, if at all, only roughly and incompletely, constitutes a minimal sufficient condition for dental decay—but not a necessary one, for there are other combinations of factors, which do not include sweet-eating, which would also make teeth decay, but which we can specify, if at all, only roughly and incompletely. That is, if ' A ' now represents sweet-eating, ' P ' dental decay, and ' F ' the class of human beings with some of their own teeth, we can say that, for some X and Y , (AX or Y) is necessary and sufficient for P in F , and we *may* be able to

go beyond this merely existentially quantified statement to at least a partial specification of the X and T in question. That is, we can say that ($A \dots$ or \dots) is a necessary and sufficient condition, but that A itself is only an *INUS* condition. And the same holds for many general causal statements of the form " A causes (or produces) P ." It is in this sense that the application of a potential difference to the ends of a copper wire produces an electric current in the wire; that a rise in the temperature of a piece of metal makes it expand; that moisture rusts steel; that exposure to various kinds of radiation causes cancer, and so on.

However, it is true that not all ordinary general causal statements are of this sort. Some of them are implicit statements of functional dependence. Functional dependence is a more complicated relationship of which necessity and sufficiency can be regarded as special cases. (It is briefly discussed in § 7 below.) Here too what we commonly single out as causing some result is only one of a number of factors which jointly affect the result. Again, some causal statements pick out something that is not only an *INUS* condition, but also a necessary condition. Thus we may say that the yellow fever virus is the cause of yellow fever. (This statement is not, as it might appear to be, tautologous, for the yellow fever virus and the disease itself can be independently specified.) In the field in question—human beings—the injection of this virus is not by itself a sufficient condition for this disease, for persons who have once recovered from yellow fever are thereafter immune to it, and other persons can be immunized against it. The injection of the virus, combined with the absence of immunity (natural or artificial), and perhaps combined with some other factors, constitutes a sufficient condition for the disease. Beside this, the injection of the virus is a necessary condition of the disease. If there is more than one complex sufficient condition for yellow fever, the injection of the virus into the patient's bloodstream (either by a mosquito or in some other way) is a factor included in every such sufficient condition. If ' A ' stands for this factor, the necessary and sufficient condition has the form ($A \dots$ or $A \dots$ etc.), where A occurs in every disjunct. We sometimes note the difference between this and the standard case by using the phrase "the cause." We may say not merely that this virus *causes* yellow fever, but that it is *the cause* of yellow fever; but we would say only that sweet-eating *causes* dental decay, not

that it is *the cause* of dental decay. But about an individual case we could say that sweet-eating was *the cause* of the decay of this person's teeth, meaning (as in § 1 above) that the only sufficient condition present here was the one of which sweet-eating is a nonredundant part. Nevertheless, there will not in general be any one item which has a unique claim to be regarded as *the cause* even of an individual event, and even after the causal field has been determined. Each of the moments in the minimal sufficient condition, or in each minimal sufficient condition, that was present can equally be regarded as the cause. They may be distinguished as predisposing causes, triggering causes, and so on, but it is quite arbitrary to pick out as "main" and "secondary," different moments which are equally nonredundant items in a minimal sufficient condition, or which are moments in two minimal sufficient conditions each of which makes the other redundant.²²

§ 4. NECESSITY AND SUFFICIENCY

One possible account of general statements of the forms " S is a necessary condition of T " and " S is a sufficient condition of T "—where ' S ' and ' T ' are general terms—is that they are equivalent to simple universal propositions. That is, the former is equivalent to "All T are S " and the latter to "All S are T ." Similarly, " S is necessary for T in the field F " would be equivalent to "All FT are S ," and " S is sufficient for T in the field F " to "All FS are T ." Whether an account of this sort is adequate is, of course, a matter of dispute; but it is not disputed that these statements about necessary and sufficient conditions at least *entail* the corresponding universals. I shall work on the assumption that this account is adequate, that general statements of necessity and sufficiency are equivalent to universals: it will be worth while to see how far this account will take us, how far we are able, in terms of it, to understand how we use, support, and criticize these statements of necessity and sufficiency.

A directly analogous account of the corresponding singular statements is not satisfactory. Thus it will not do to say that "A short-circuit here was a necessary condition of a fire in this house" is equivalent to "All cases of this house's catching fire are cases of a short-circuit occurring here," because the latter is automatically true if this house has caught fire only once and a short-circuit has

²² Cf. Marc-Wogau's concluding remarks, *op. cit.*, pp. 232-233.

occurred on that occasion, but this is not enough to establish the statement that the short-circuit was a necessary condition of the fire; and there would be an exactly parallel objection to a similar statement about a sufficient condition.

It is much more plausible to relate singular statements about necessity and sufficiency to certain kinds of non-material conditionals. Thus "A short-circuit here was a necessary condition of a fire in this house" is closely related to the counterfactual conditional "If a short-circuit had not occurred here this house would not have caught fire," and "A short-circuit here was a sufficient condition of a fire in this house" is closely related to what Goodman has called the factual conditional, "Since a short-circuit occurred here, this house caught fire."

However, a further account would still have to be given of these non-material conditionals themselves. I have argued elsewhere²³ that they are best considered as condensed or telescoped arguments, but that the statements used as premisses in these arguments are no more than simple factual universals. To use the above-quoted counterfactual conditional is, in effect, to run through an incomplete argument: "Suppose that a short-circuit did not occur here, then the house did not catch fire." To use the factual conditional is, in effect, to run through a similar incomplete argument, "A short-circuit occurred here; therefore the house caught fire." In each case the argument might in principle be completed by the insertion of other premisses which, together with the stated premiss, would entail the stated conclusion. Such additional premisses may be said to *sustain* the non-material conditional. It is an important point that someone can use a non-material conditional without completing or being able to complete the argument, without being prepared explicitly to assert premisses that would sustain it, and similarly that we can understand such a conditional without knowing exactly how the argument would or could be completed. But to say that a short-circuit here was a necessary condition of a fire in this house is to say that there is some set of true propositions which would sustain the above-stated counterfactual, and to say that it was a sufficient condition is to say

that there is some set of true propositions which would sustain the above-stated factual conditional. If this is conceded, then the relating of singular statements about necessity and sufficiency to non-material conditionals leads back to the view that they refer indirectly to certain simple universal propositions. Thus if we said that a short-circuit here was a necessary condition for a fire in this house, we should be saying that there are true universal propositions from which, together with true statements about the characteristics of this house, and together with the supposition that a short-circuit did not occur here, it would follow that the house did not catch fire. From this we could infer the universal proposition which is the more obvious, but unsatisfactory, candidate for the analysis of this statement of necessity, "All cases of this house's catching fire are cases of a short-circuit occurring here," or, in our symbols, "All FP are A ." We can use this to represent approximately the statement of necessity, on the understanding that it is to be a consequence of some set of wider universal propositions, and is not to be automatically true merely because there is only this one case of an FP , of this house's catching fire.²⁴ A statement that A was a sufficient condition may be similarly represented by "All FA are P ." Correspondingly, if all that we want to say is that (A ... or ...) was necessary and sufficient for P in F , this will be represented approximately by the pair of universals "All FP are (A ... or ...) and all F (A ... or ...) are P ," and more accurately by the statement that there is some set of wider universal propositions from which, together with true statements about the features of F , this pair of universals follows. This, therefore, is the fuller analysis of the claim that in a particular case A is an *INUS* condition of P in F , and hence of the singular statement that A caused P . (The statement that A is *at least* an *INUS* condition includes other alternatives, corresponding to cases where the necessary and sufficient condition is (A or ...), A ..., or A .)

Let us go back now to general statements of necessity and sufficiency and take F as a class, not as an individual. On the view that I am adopting, at least provisionally, the statement that Z is a necessary and sufficient condition for P in F is

²³ "Counterfactuals and Causal Laws," *Analytical Philosophy*, ed. R. J. Butler (Oxford, 1962), pp. 66-80.

²⁴ This restriction may be compared with one which Nagel imposes on laws of nature: "the vacuous truth of an unrestricted universal is not sufficient for counting it a law; it counts as a law only if there is a set of other assumed laws from which the universal is logically derivable" (*The Structure of Science* [New York, 1961], p. 60). It might have been better if he had added "or if there is some other way in which it is supported (ultimately) by empirical evidence." Cf. my remarks in "Counterfactuals and Causal Laws," *op. cit.*, pp. 72-74, 78-80.

equivalent to "All FP are Z and all FZ are P ." Similarly, if we cannot completely specify a necessary and sufficient condition for P in F , but can only say that the formula " $(A... \text{ or } ...)$ " represents such a condition, this is equivalent to the pair of incomplete universals, "All FP are $(A... \text{ or } ...)$ and all $F(A... \text{ or } ...) \text{ are } P$." In saying that our general causal statements often do no more than specify an INUS condition, I am therefore saying that much of our ordinary causal knowledge is knowledge of such pairs of incomplete universals, of what we may call elliptical or *gappy* causal laws.

§ 5. EVIDENCE FOR CAUSAL CONNECTIONS

If we assume that the general causal statement that A causes P , or the singular causal statement that A caused P , often makes the claims set out in §§ 1, 2, 3, and 4, including the claim that A is at least an INUS condition of P , then we can give an account of a combination of reasoning and observation which constitutes evidence for these causal statements.

This account is based on what von Wright calls a complex case²⁵ of the Method of Difference. Like any other method of eliminative induction, this can be formulated in terms of an assumption, an observation, and a conclusion which follows by a deductively valid argument from the assumption and the observation together. To get any positive conclusion by a process of elimination, we must assume that the result (the phenomenon a cause of which we are going to discover) has *some* cause in the sense that there is some condition the occurrence of which is both necessary and sufficient for the occurrence (as a rule, shortly afterwards) of the result. Also, if we are to get anywhere by elimination, we must assume that the range of possibly relevant causal factors, the items that might in some way constitute this necessary and sufficient condition, is restricted in some way. On the other hand, even if we had specified some such set of possibly relevant factors, it would in most cases be quite implausible to assume that the supposed necessary and sufficient condition is identical with just one of these factors on its own, and fortunately we have no need to do so. If we represent each possibly relevant factor as a single term, the natural assumption to make is merely

that the supposed necessary and sufficient condition will be represented by a formula which is constructed in some way out of some selection of these single terms, by means of negation, conjunction, and disjunction. However, any formula so constructed is equivalent to some formula in disjunctive normal form—that is, one in which negation, if it occurs, is applied only to single terms, and conjunction, if it occurs, only to single terms and/or negations of single terms. So we can assume without loss of generality that the formula of the supposed necessary and sufficient condition is in disjunctive normal form, that it is at most a disjunction of conjunctions in which each conjunct is a single term or the negation of one, that is, a formula such as " $(ABC \text{ or } GH \text{ or } J)$." Summing this up, the assumption that we require will have this form:

For some Z , Z is a necessary and sufficient condition for the phenomenon P in the field F , that is, all FP are Z and all FZ are P , and Z is a condition represented by some formula in disjunctive normal form all of whose constituents are taken from the range of possibly relevant factors A, B, C, D, E , etc.

Along with this assumption, we need an observation which has the form of the classical difference observation described by Mill. This we can formulate as follows:

There is an instance I_1 , in which P occurs, and there is a negative case N_1 , in which P does not occur, such that one of the possibly relevant factors (or the negation of one), say A , is present in I_1 and absent from N_1 , but each of the other possibly relevant factors is either present in both I_1 and N_1 or absent both from I_1 and from N_1 .

We can set out an example of such an observation as follows, using ' a ' and ' p ' to stand for "absent" and "present."

	P	A	B	C	D	E	
I_1	p	p	p	a	a	p	} etc.
N_1	a	a	p	a	a	p	

Given the above-stated assumption, we can reason in the following way about any such observation:

²⁵ *A Treatise on Induction and Probability* (New York, 1951), pp. 90 ff. The account that I am here giving of the Method of Difference, and that I would give of the eliminative methods of induction in general, differs, however, in several respects from that of von Wright. An article on "Eliminative Methods of Induction," which sets out my account, is to appear in the *Encyclopedia of Philosophy*, edited by Paul Edwards, to be published by the Free Press of Glencoe, Collier-Macmillan.

Since P is absent from N_1 , every sufficient condition for P is absent from N_1 , and therefore every disjunct in Z is absent from N_1 . Every disjunct in Z which does not contain A is therefore also absent from I_1 . But since P is present in I_1 , and Z is a necessary condition for P , Z is present in I_1 . Therefore at least one disjunct in Z is present in I_1 . Therefore at least one disjunct in Z contains A .

What this shows is that Z , the supposed necessary and sufficient condition for P in F , is either A itself, or a conjunction containing A , or a disjunction containing as a disjunct either A itself or a conjunction containing A . That is, Z has one of these four forms: A ; $A...$; $(A \text{ or } ...)$; $(A... \text{ or } ...)$. We can sum these up by saying that Z has the form $(A--- \text{ or } ---)$, where the dashes indicate that these parts of the formula may or may not be filled in. This represents briefly the statement that A is at least an *INUS* condition. It follows also that if there are in the (unknown) formula which represents the complete necessary and sufficient condition any disjuncts not containing A , none of them was present as a whole in N_1 (but of course some of their component terms may have been present there), and also that in at least one of the disjuncts that contains A , the terms, if any, conjoined with A stand for factors (or negations of factors) that were present in I_1 . This is all that follows from this single observation. But in general other observations will show that the dotted spaces do need to be filled in, and that A alone is neither sufficient nor necessary for P in F . We can then infer that the necessary and sufficient condition actually has the form $(A... \text{ or } ...)$, and that A itself is only an *INUS* condition.

This analysis is so far merely formal, and we have still to consider whether such a method can be, or is, actually used, whether an assumption of the sort required can be justified and whether an observation of the sort required can ever be made. Even at this stage, however, it is worth noting that the Method of Difference does not require the utterly unrealistic sort of assumption used in what von Wright calls the simple case—namely, that the supposed necessary and sufficient condition is some single factor on its own—but that the much less restrictive assumption used here will still yield information when it is combined with nothing more than the classical difference observation. It is worth

noting also that the information thus obtained, though it falls far short of what von Wright calls absolutely perfect analogy, that is, of a full specification of a necessary and sufficient condition, is information of exactly the form that is implicit in our ordinary causal assertions, both singular and general.²⁶

But can observations of the kind required be made? A preliminary answer is that the typical controlled experiment is an attempt to approximate to an observation of this sort. The experimental case corresponds to our I_1 , the control case to our N_1 , and the experimenter tries to insure that there will be no possibly relevant difference between these two except the one whose effect he is trying to determine, our A . Any differential outcome, present in the experimental case but not in the control case, is what he takes to be this effect, corresponding to our P .

The before-and-after observation is a particularly important variety of this kind. Suppose, for example, that we take a piece of blue litmus paper and dip it in a certain liquid, and it turns red. The situation before it is dipped provides the negative case N_1 ; the situation after it is dipped provides the instance I_1 . As far as we can see, no other possibly relevant feature of the situation has changed, so that I_1 and N_1 are alike with regard to all possibly relevant factors except A , the paper's being dipped in a liquid of this sort, but the result P , the paper's turning red, is present in I_1 but not in N_1 . We can take this in either of two ways. First, we may take the field F to be pieces of blue litmus paper, and if we assume that in this field there is some necessary and sufficient condition for P , made up in some way from some selection from the factors we are considering as possibly relevant, we can conclude that $(A--- \text{ or } ---)$ is necessary and sufficient for P in F . Other observations may show that A alone is neither necessary nor sufficient, and hence that the necessary and sufficient condition is $(A... \text{ or } ...)$. Thus we can establish the gappy causal law, "All FP are $(A... \text{ or } ...)$ and all $F(A... \text{ or } ...)$ are P ." This amounts to the assertion that in some circumstances being dipped in a liquid of this sort turns blue litmus paper red. Secondly, we can take the field (which we shall here call F_1) to be this particular piece of paper, and what the experiment then establishes

²⁶ What is established by the present method may be compared with the four claims listed in § 1 above, that A is at least an *INUS* condition, that A was present on the occasion in question, that the factors represented by 'X'—that is, the other moments in at least one minimal sufficient condition in which A is a moment—were present, and that every disjunct in Z which does not contain A —that is, every minimal sufficient condition which does not contain A —was absent.

is the singular causal statement that on this particular occasion the dipping in this liquid turned this piece of paper red. This is established in accordance with the analysis of singular causal statements completed in § 4. For the experiment, together with the assumption, has established the wider universals indicated by the above-stated gappy causal law. It has shown that for some X and Y all FP are $(AX \text{ or } Y)$ and all $F(AX \text{ or } Y)$ are P , and from these, since F_1 is an F (that is, this piece of paper is a piece of blue litmus paper), it follows that for some X and Y all F_1P are $(AX \text{ or } Y)$ and all $F_1(AX \text{ or } Y)$ are P . Also, ' X ' represents circumstances which were present on this occasion, and ' Y ' circumstances which were not present in N_1 , the "before" situation. That is to say, the observation, together with the appropriate assumption, entails that there are true propositions which sustain the counterfactual and factual conditionals, "If, in the circumstances, this paper had not been dipped in this liquid it would not have turned red, but since it was dipped it did turn red"; but it does not fully determine what these propositions are, it does not fill in the gaps in the causal laws which sustain these conditionals. The importance of this is that it shows how an observation can reveal not merely a sequence but a causal sequence: what we discover is not merely that the litmus paper was dipped *and then* turned red, but that the dipping *made* it turn red.

It is worth noting that despite the stress traditionally laid, in accounts of the Method of Difference, on the requirement that there should be only *one* point of difference between I_1 and N_1 , very little really turns upon this. For suppose that two of our possibly relevant factors, say A and B , were both present in I_1 and both absent from N_1 , but that each of the other possibly relevant factors was either present in both or absent from both. Then reasoning parallel to that given above will show that at least one of the disjuncts in Z either contains A or contains B (and may contain both). That is, this observation still serves to show that the cluster of factors (A, B) contains something that is at least an INUS condition of P in F , whether this condition turns out in the end to be A alone, or B alone, or the conjunction AB , or the disjunction $(A \text{ or } B)$. And similar considerations apply if there are more than two points of difference between I_1 and N_1 . However many there are, an observation of this form, coupled with our assumption, shows that a cause in our sense (in general an INUS condition) lies somewhere within the cluster of terms,

positive or negative, in respect of which I_1 differs from N_1 . (Note that it does *not* show that the other terms, those common to I_1 and N_1 , are causally irrelevant; our reasoning does not exclude factors as irrelevant, but positively locates some of the relevant factors within the differentiating cluster.)

This fact rebuts the criticism sometimes leveled against the eliminative methods that they presuppose and require a finally satisfactory analysis of causal factors into their simple components, which we never actually achieve. On the contrary, any distinction of factors, however rough, enables us to start using such a method. We can proceed, and there is no doubt that discovery has often proceeded, by what we may call the *progressive localization of a cause*. Using the Method of Difference in a very rough way, we can discover first, say, that the drinking of wine causes intoxication. That is, the cluster of factors which is crudely summed up in the single term "the drinking of wine" contains somewhere within it an INUS condition of intoxication; and we can subsequently go on to distinguish various possibly relevant factors within this cluster, and by further observations of the same sort locate a cause of intoxication more precisely. In a context in which this cluster is either introduced or excluded as a whole, it is correct to say that the introduction of this cluster was non-redundant or necessary *post factum*, and experiments can establish this, even if, in a different context, in which distinct items in the cluster are introduced or excluded separately, it would be correct to say that only one item, the alcohol, was nonredundant or necessary *post factum*, and this could be established by more exact experimentation.

One merit of this formal analysis is that it shows in what sense a method of eliminative induction, such as the Method of Difference, rests upon a deterministic principle or presupposes the uniformity of nature. In fact, each application of this method requires an assumption which in one respect says much less than this, in another a little more. No sweeping general assumption is needed: we need not assume that every event has a cause, but merely that for events of the kind in question, P , in the field in question, F , there is some necessary and sufficient condition. But—and this is where we need something more than determinism or uniformity in general—we must also assume that this condition is constituted in some way by some selection from a restricted range of possibly relevant factors.

It is this further assumption that raises a doubt

about the use of this method to make causal discoveries. As for the mere deterministic assumption that the phenomenon in question has some necessary and sufficient condition, we may be content to say that this is one which we simply do make in all inquiries of this kind, and leave its justification to be provided by whatever solution we can eventually find for the general problem of induction. But the choice of a range of possibly relevant factors cannot be brushed aside so easily. Also, the wider a range of possibly relevant factors we admit, the harder it will be to defend the claim that I_1 and N_1 are observed to be alike with respect to all the possibly relevant factors except the one, or the indicated cluster of factors, in which they are observed to differ. Alternatively, the more narrowly the range of possibly relevant factors is restricted, the easier it will be to defend the claim that we have made an observation of the required form, but at the same time the less plausible will our assumption be.

However, this difficulty becomes less formidable if we consider the assumption and the observation together. We want to be able to say that there is no possibly relevant difference, other than the one (or ones) noted, between I_1 and N_1 . We need not draw up a complete list of possibly relevant factors before we make the observation. In practice we usually assume that a causally relevant factor will be in the spatial neighborhood of the instance of the field in or to which the effect occurs in I_1 , or fails to occur in N_1 , and it will either occur shortly before or persist throughout the time at which the effect occurs in I_1 , or might have occurred, but did not, in N_1 . No doubt in a more advanced application of the Method of Difference within an already-developed body of causal knowledge we

can restrict the range of possibly relevant factors much more narrowly and can take deliberate steps to exclude interferences from our experiments; but I am suggesting that even our most elementary and primitive causal knowledge rests upon implicit applications of this method, and the spatio-temporal method of restricting possibly relevant factors is the only one initially available. And perhaps it is all we need. Certainly in terms of it the observer could say, about the litmus paper, for example, "I cannot see any difference, other than the dipping into this liquid, between the situation in which the paper turned red and that in which it did not, that might be relevant to this change."

It may be instructive to compare the Method of Difference as a logical ideal with any actual application of it. If the assumption and the observation were known to be true, then the causal conclusion would be established. Consequently, anything that tells in favor of both the assumption and the observation tells equally in favor of the causal conclusion. No doubt we are never in a position to say that they are known to be true, and therefore that the conclusion is established; but we are often in a position to say that, given the deterministic part of the assumption, we cannot see any respect in which they are not true (since we cannot see any difference that might be relevant between I_1 and N_1), and consequently that we cannot see any escape from the causal conclusion. In this sense at least we can say that an application of this method confirms a causal conclusion: the observer has looked for but failed to find an escape from this conclusion.²⁷

In practice we do not rely as much on single observations as this account might suggest. We assure ourselves that it was the dipping in this

²⁷ An account of how eliminative inductive reasoning supports causal conclusions is given by J. R. Lucas in the article cited in n. 16 above. His account differs from mine in many details, but agrees with it in general outline. Contrast with this the remarks of von Wright, *op. cit.*, p. 135: "... in normal scientific practice we have to reckon with plurality rather than singularity, and with complexity rather than simplicity of conditions. This means that the weaker form of the Deterministic Postulate, or the form which may be viewed as a reasonable approximation to what is commonly known as the Law of Universal Causation, is practically useless as a supplementary premiss or 'presupposition' of induction." I hope I have shown that this last remark is misleading.

It has been argued by A. Michotte (*La perception de la causalité* [Louvain, 1946], translated by T. R. and E. Miles as *The Perception of Causality* [London, 1963]) that we have in certain cases an immediate perception or impression of causation. His two basic experimental cases are these. In one, an object A approaches another object B ; on reaching B , A stops and B begins to move off in the same direction; here the observer gets the impression that A has "launched" B , has set B in motion. In the other case, A continues to move on reaching B , and B moves at the same speed and in the same direction; here the observer gets the impression that A is carrying B with it. In both cases observers typically report that A has caused the movement of B . Michotte argues that it is an essential feature of observations that give rise to this causal impression that there should be two distinguishable movements, that of the "agent" A and that of the "patient" B , but also that it is essential that the movement of the patient should in some degree copy or duplicate that of the agent.

This would appear to be a radically different account of the way in which we can detect causation by observing a single sequence, for on Michotte's view our awareness of causation can be direct, perceptual, and non-inferential. It must be conceded that not only spatio-temporal continuity, but also qualitative continuity between cause and effect (*l'ampliation du*

liquid that turned the litmus paper red by dipping other pieces of litmus paper and seeing them, too, turn red just after they are dipped. This repetition is effective because it serves as a check on the possibility that some other relevant change might have occurred, unnoticed, just at the moment when the first piece of litmus paper was dipped in the liquid. After a few trials it will be most unlikely that any other relevant change has kept on occurring just as each piece was dipped (or even that there has been a succession of different relevant changes at the right times). Of course, it may be that there is some other relevant change (or set of relevant changes) which keeps on occurring just as each paper is dipped because it is linked with the dipping by what Mill calls "some fact of causation."²⁸ If so, then this other relevant change may be regarded as part of a cluster of factors which can be grouped together under the title "the dipping of the paper in this liquid," taking this in a broad sense, as possibly including items other than the actual entry of the paper into the liquid. But if this is not so, then it would be a sheer coincidence if this other relevant change kept on occurring just as each piece of paper was dipped, or if there was a succession of relevant changes at the right times. The hypothesis that such coincidences have continued will soon become implausible, even if it cannot be conclusively falsified.²⁹ It is an important point that it is not the repetition as such that supports the conclusion that the dipping causes the turning red, but the repetition of a sequence which, on each single occasion, is already *prima facie* a causal one. The repetition tends to disconfirm the set of hypotheses each of which explains a single sequence of a dipping followed by a turning red as a mere

coincidence, and by contrast it confirms the hypothesis that in each such single sequence the dipping is causally connected with the change of color.

The analysis offered here of the Method of Difference has this curious consequence: in employing this method we are liable to use the word "cause" in different senses at different stages. In the assumption, it is said that the phenomenon *P* has some "cause," meaning some necessary and sufficient condition; but the "cause" actually found—*A* in our formal example—may be only an INUS condition. But we do need to assume that *something* is both necessary and sufficient for *P* in *F* to be able to conclude that *A* is at least an INUS condition, that it is a moment in a minimal sufficient condition that was present, and that it was necessary *post factum*.

§ 6. FALSIFICATION OF INCOMPLETE STATEMENTS

A possible objection to this account is that the gappy laws and singular statements used here are so incomplete that they are internally guaranteed against falsification and are therefore not genuine scientific statements at all. However, it is not a satisfactory criterion of a scientific statement that it should be exposed to conclusive falsification: what is important is that to treat a statement as a scientific hypothesis involves handling it in such a way that evidence would be allowed to tell against it. And there are ways in which evidence can be, and is, allowed to tell against a statement which asserts that something is an INUS condition.

Suppose, for example, that by using the I_1 and N_1 set out in § 5 above we have concluded that *A* is at least an INUS condition of *P*—taking this both as a singular causal statement about an individual

mouvement), are important ingredients in the primitive concept of causation; they may contribute to the notion of causal "necessity"; and both these continuities can sometimes be directly perceived. But it is equally clear that these continuities are not in general required either as observed or as postulated features of a causal sequence, and that a sequence which has these continuities may fail to be causal. What is perceived in Michotte's examples is neither necessary nor sufficient for causal relationship as we now understand it, though it may have played an important part in the genesis of the causal concept. It is worth noting that these examples also exhibit the features stressed in my account. They present the observer with an apparently simple and isolated causal field, within which there occurs a marked change, *B*'s beginning to move. The approach of *A* is the only observed possibly relevant difference between the times when *B* is stationary and when *B* begins to move. If *B*'s beginning to move has a cause, then *A*'s approach is a suitable candidate, and nothing else that the observer is allowed to see or encouraged to suspect is so. Thus these examples could also give rise to an inferential awareness of causation, though it is true that other examples which would do this equally well would fail, and in Michotte's experiments do fail, to produce a direct impression of causation.

²⁸ E.g., in the Fifth Canon, *A System of Logic*, Book III, Chapter VIII, § 6.

²⁹ Cf. J. R. Lucas, *op. cit.*, p. 53: "It might be that two quite independent processes were going on, and we were getting constant concomitance for no reason except the chance fact that the two processes happened to keep in step. If this be so, an arbitrary disturbance in one will reveal the independence of the other. If an arbitrary disturbance in the one is followed by a corresponding alteration in the other, it always could be that it was a genuine coincidence. . . . But to argue this persistently is to make the same illicit extension of 'coincidence' as some phenomenologists do of 'illusion'. . . . It is no longer a practical possibility that we are eliminating but a Cartesian doubt."

field F_1 and as an incomplete law about the general field F . Now suppose that closer examination shows that some other factor, previously unnoticed, say K , was present in I_1 and absent from N_1 , and that we also discover (or construct experimentally) further cases I_2 and N_2 , such that the observational evidence is now of this form:

	P	A	B	C	D	E	$\dots K$	\dots
I_1	p	p	p	a	a	p	$\dots p$	\dots
N_1	a	a	p	a	a	p	$\dots a$	\dots
I_2	p	a	p	a	a	p	$\dots p$	\dots
N_2	a	p	p	a	a	p	$\dots a$	\dots

Here N_2 shows that for any X which does not contain K , AX is not sufficient: so X must contain K . But any X that contains K is present in I_2 , and *may* therefore be sufficient for P on its own, without A . This evidence does not conclusively falsify the hypothesis that A is an *INUS* condition as stated above, but it takes away all the reason that the previous evidence gave us for this conclusion. Observations of this pattern would tell against this conclusion, and would lead us to replace the view that A causes P , and caused P in I_1 , with the view that K causes P , and caused P both in I_2 and in I_1 , with A not even forming an indispensable part of the sufficient condition which was present in I_1 . (A fuller treatment of this kind of additional evidence would require accounts of the Method of Agreement and of the Joint Method, parallel to that of the Method of Difference given in § 5.)

It remains true that some of the claims made by singular causal statements and by causal laws as here analyzed—that is, claims that some factor is at least an *INUS* condition of the effect—are not conclusively falsifiable. But ordinary causal laws and singular causal statements are not conclusively falsifiable, as direct consideration will show. It is a merit of the account offered here, not a difficulty for it, that it reproduces this feature of ordinary causal knowledge.³⁰

§ 7. FUNCTIONAL DEPENDENCE AND CONCOMITANT VARIATION

As I mentioned in § 3, causal statements sometimes refer not to relations of necessity and sufficiency, nor to any more complex relations based on these, like that of being an *INUS* condition, but to relations of functional dependence. That is, the effect and the possible causal factors are things which can vary in magnitude, and the cause of

some effect P is that on whose magnitude the magnitude of P functionally depends. But causal statements of this sort can be expanded and analyzed in an account parallel to that which we have given of causal statements of the previous kinds. Again we speak of a field, individual or general, in relation to which a certain functional dependence holds. Also, we can speak of the *total cause*, the complete set of factors on whose magnitude the magnitude of P , given the field F , wholly depends: that is, variations of P in F are completely covered by a formula which is a function of the magnitudes of all of the factors in this "complete set," and of these alone. This total cause is analogous to a necessary and sufficient condition. It can be distinguished from each of the factors that compose it, each of which is causally relevant to the effect, but it is not the whole cause of its variations: each of these *partial causes* is analogous to an *INUS* condition.

The problem of finding a cause in this new sense would require, for its full solution, the completion of two tasks. We should have both to identify all the factors in this total cause, and also to discover in what way the effect depends upon them—that is, to discover the law of functional dependence of the effect on the total cause, or the partial differential equations relating it to each of the partial causes. The first—but only the first—of these two tasks can be performed by what is really the Method of Concomitant Variation, developed in a style analogous to that in which the Method of Difference was developed in § 5. That is, we assume that there is something on which the magnitude of P in F functionally depends, and that there is a restricted set of possibly relevant factors; then if while all other possibly relevant factors are held constant one factor, say A , varies and P also varies, it follows that A is at least a partial cause, that it is one of the actually relevant factors. It is this relationship that is commonly asserted by statements of such forms as " A affects P " and "On this occasion A affected P ." Some of our causal statements, singular or general, have just this force, and all that I am trying to show here is that these statements can be supported by reasoning along the lines of the Method of Concomitant Variation, developed analogously with the development in § 5 of the Method of Difference. Just as we there assumed that there was some necessary and sufficient condition, and by combining this assumption with our observations dis-

³⁰ This was pointed out by D. C. Stove.

covered something which is at least an INUS condition, so we here assume that there is some total cause and so discover something which is at least a partial cause. However, a complete account of the Method of Concomitant Variation would involve the examination of several other cases besides the one sketched here.³¹ For our present purpose, we need note only that there is this functional dependence part of the concept of causation as well as the presence-or-absence part, indeed that the latter can be considered as a special limiting case of the former,³² but that the two parts are systematically analogous to one another, and that our knowledge of both singular and general causal relationships of these two kinds can be accounted for on corresponding principles.

§ 8. THE DIRECTION OF CAUSATION

This account of causation is still incomplete, in that nothing has yet been said about the direction of causation, about what distinguishes *A* causing *P* from *P* causing *A*. This is a difficult question, and it is linked with the equally difficult question of the direction of time. I cannot hope to resolve it completely here, but I shall state some of the relevant considerations.³³

First, it seems that there is a relation which may be called *causal priority*, and that part of what is meant by "*A* caused *P*" is that this relation holds in one direction between *A* and *P*, not the other. Secondly, this relation is not identical with temporal priority; it is conceivable that there should be evidence for a case of backward causation, for *A* being causally prior to *P* whereas *P* was temporally prior to *A*. Most of us believe, and I think with good reason, that backward causation does not occur, so that we can and do normally use temporal order to limit the possibilities about causal order; but the connection between the two is synthetic. Thirdly, it could be objected to the analysis of "necessary" and "sufficient" offered in § 4 above that it omits any reference to causal order, whereas our most common use of "necessary" and "sufficient" in causal contexts includes such a reference. Thus "*A* is (causally) sufficient for *B*" says "If *A*, then *B*, and *A* is causally prior to *B*," but "*B* is (causally) necessary for *A*" is not equivalent to

this: it says "If *A*, then *B*, and *B* is causally prior to *A*." However, it is simpler to use "necessary" and "sufficient" in senses which exclude this causal priority, and to introduce the assertion of priority separately into our accounts of "*A* caused *P*" and "*A* causes *P*." Fourthly, although "*A* is (at least) an INUS condition of *P*" is not synonymous with "*P* is (at least) an INUS condition of *A*," this difference of meaning cannot exhaust the relation of causal priority. If it did exhaust it, the direction of causation would be a trivial matter, for, given that there is some necessary and sufficient condition of *A* in the field, it can be proved that if *A* is (at least) an INUS condition of *P*, then *P* is also (at least) an INUS condition of *A*: we can construct a minimal sufficient condition of *A* in which *P* is a moment.³⁴

Fifthly, it is often suggested that the direction of causation is linked with controllability. If there is a causal relation between *A* and *B*, and we can control *A* without making use of *B* to do so, and the relation between *A* and *B* still holds, then we decide that *B* is not causally prior to *A* and, in general, that *A* is causally prior to *B*. But this means only that if one case of causal priority is known, we can use it to determine others: our rejection of the possibility that *B* is causally prior to *A* rests on our knowledge that our action is causally prior to *A*, and the question how we know the latter, and even the question of what causal priority is, have still to be answered. Similarly, if one of the causally related kinds of event, say *A*, can be randomized, so that occurrences of *A* are either not caused at all, or are caused by something which enters this causal field *only* in this way, by causing *A*, we can reject both the possibility that *B* is causally prior to *A* and the possibility that some common cause is prior both to *A* and separately to *B*, and we can again conclude that *A* is causally prior to *B*. But this still means only that we can infer causal priority in one place if we first know that it is absent from another place. It is true that our knowledge of the direction of causation in ordinary cases is thus based on what we find to be controllable, and on what we either find to be random or find that we can randomize; but this cannot without circularity be taken as providing a full account either of what we mean.

³¹ I have given a fuller account of this method in the article cited in n. 25.

³² Cf. J. R. Lucas, *op. cit.*, p. 65.

³³ As was mentioned in n. 21, Scriven's fifth difficulty and refinement are concerned with this point (*op. cit.*, pp. 411-412), but his answer seems to me inadequate. Lucas touches on it (*op. cit.*, pp. 51-53). The problem of temporal asymmetry is discussed, e.g., by J. J. C. Smart, *Philosophy and Scientific Realism* (London, 1963), pp. 142-148, and by A. Grünbaum in the article cited in n. 36 below.

³⁴ I am indebted to one of the referees for correcting an inaccurate statement on this point in an earlier version.

by causal priority or of how we know about it.

A suggestion put forward by Popper about the direction of time seems to be relevant here.³⁵ If a stone is dropped into a pool, the entry of the stone will explain the expanding circular waves. But the reverse process, with contracting circular waves, "would demand a vast number of distant coherent generators of waves the coherence of which, to be explicable, would have to be shown . . . as originating from one centre." That is, if *B* is an occurrence which involves a certain sort of "coherence" between a large number of separated items, whereas *A* is a single event, and *A* and *B* are causally connected, *A* will explain *B* in a way in which *B* will not explain *A* unless some other single event, say *C*, first explains the coherence in *B*. Such examples give us a *direction of explanation*, and it may be that this is the basis, or part of the basis, of the relation I have called causal priority.

§ 9. CONCLUSIONS

Even if Mill was wrong in thinking that science consists mainly of causal knowledge, it can hardly be denied that such knowledge is an indispensable element in science, and that it is worth while to investigate the meaning of causal statements and the ways in which we can arrive at causal knowledge. General causal relationships are among the items which a more advanced kind of scientific theory explains, and is confirmed by its success in explaining. Singular causal assertions are involved in almost every report of an experiment: doing such and such *produced* such and such an effect. Materials are commonly identified by their causal properties: to recognize something as a piece of a certain material, therefore, we must establish singular causal assertions about it, that this object affected that other one, or was affected by it, in such and such a way. Causal assertions are embedded in both the results and the procedures of scientific investigation.

The account that I have offered of the force of various kinds of causal statements agrees both with our informal understanding of them and with

accounts put forward by other writers: at the same time it is formal enough to show how such statements can be supported by observations and experiments, and thus to throw a new light on philosophical questions about the nature of causation and causal explanation and the status of causal knowledge.

One important point is that, leaving aside the question of the direction of causation, the analysis has been given entirely within the limits of what can still be called a regularity theory of causation, in that the causal laws involved in it are no more than straightforward universal propositions, although their terms may be complex and perhaps incompletely specified. Despite this limitation, I have been able to give an account of the meaning of statements about singular causal sequences, regardless of whether such a sequence is or is not of a kind that frequently recurs: repetition is not essential for causal relation, and regularity does not here disappear into the mere fact that this single sequence has occurred. It has, indeed, often been recognized that the regularity theory could cope with single sequences if, say, a unique sequence could be explained as the resultant of a number of laws each of which was exemplified in many other sequences; but my account shows how a singular causal statement can be interpreted, and how the corresponding sequence can be shown to be causal, even if the corresponding complete laws are not known. It shows how even a unique sequence can be directly recognized as causal.

One consequence of this is that it now becomes possible to reconcile what have appeared to be conflicting views about the nature of historical explanation. We are accustomed to contrast the "covering-law" theory adopted by Hempel, Popper, and others with the views of such critics as Dray and Scriven who have argued that explanations and causal statements in history cannot be thus assimilated to the patterns accepted in the physical sciences.³⁶ But while my basic analysis of singular causal statements in §§ 1 and 2 agrees closely with Scriven's, I have argued in § 4 that this analysis can be developed in terms of complex

³⁵ "The Arrow of Time," *Nature*, vol. 177 (1956), p. 538; also vol. 178, p. 382 and vol. 179, p. 1297.

³⁶ See, for example, C. G. Hempel, "The Function of General Laws in History," *Journal of Philosophy*, vol. 39 (1942), reprinted in *Readings in Philosophical Analysis*, ed. by H. Feigl and W. Sellars (New York, 1949), pp. 459-471; C. G. Hempel and P. Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science*, vol. 15 (1948), reprinted in *Readings in the Philosophy of Science*, ed. by H. Feigl and M. Brodbeck (New York, 1953), pp. 319-352; K. R. Popper, *Logik der Forschung* (Vienna, 1934), translation *The Logic of Scientific Discovery* (London, 1959), pp. 59-60, also *The Open Society* (London, 1952), vol. II, p. 262; W. Dray, *Laws and Explanation in History* (Oxford, 1957); N. Rescher, "On Prediction and Explanation," *British Journal for the Philosophy of Science*, vol. 9 (1958), pp. 281-290; various papers in *Minnesota Studies in the Philosophy of Science*, vol. III, ed. by H. Feigl and G. Maxwell (Minneapolis, 1962); A. Grünbaum, "Temporally-asymmetric Principles,

and elliptical universal propositions, and this means that wherever we have a singular causal statement we shall still have a covering law, albeit a complex and perhaps elliptical one. Also, I have shown in § 5, and indicated briefly, for the functional dependence variants, in § 7, that the evidence which supports singular causal statements also supports general causal statements or covering laws, though again only complex and elliptical ones. Hempel recognized long ago that historical accounts can be interpreted as giving incomplete "explanation sketches," rather than what he would regard as full explanations, which would require fully-stated covering laws, and that such sketches are also common outside history. But in these terms what I am saying is that explanation sketches and the related elliptical laws are often all that we can discover, that they play a part in all sciences, that they can be supported and even established without being completed, and do not serve merely as preliminaries to or summaries of complete deductive explanations. If we modify the notion of a covering law to admit laws which not only are complex but also are known only in an elliptical form, the covering-law theory can accommodate many of the points that have been made in criticism of it, while preserving the structural similarity of explanation in history and in the physical sciences. In this controversy, one point

at issue has been the symmetry of explanation and prediction, and my account may help to resolve this dispute. It shows, in agreement with what Scriven has argued, how the actual occurrence of an event in the observed circumstances—the I_1 of my formal account in § 5—may be a vital part of the evidence which supports an explanation of that event, which shows that it was A that caused P on this occasion. A prediction on the other hand cannot rest on observation of the event predicted. Also, the gappy law which is sufficient for an explanation will not suffice for a prediction (or for a retrodiction): a statement of initial conditions together with a gappy law will not entail the assertion that a specific result will occur, though of course such a law may be, and often is, used to make tentative predictions the failure of which will not necessarily tell against the law. But the recognition of these differences between prediction and explanation does not affect the covering-law theory as modified by the recognition of elliptical laws.

Although what I have given is primarily an account of physical causation, it may be indirectly relevant to the understanding of human action and mental causation. It is sometimes suggested that our ability to recognize a single occurrence as an instance of mental causation is a feature which distinguishes mental causation from physical or "Humean" causation.³⁷ But this suggestion arises

Parity between Explanation and Prediction, and Mechanism versus Teleology," *Philosophy of Science*, vol. 29 (1962), pp. 146–170.

Dray's criticisms of the covering-law theory include the following: we cannot state the law used in an historical explanation without making it so vague as to be vacuous (*op. cit.*, especially pp. 24–37) or so complex that it covers only a single case and is trivial on that account (p. 99); the historian does not come to the task of explaining an event with a sufficient stock of laws already formulated and empirically validated (pp. 42–43); historians do not need to replace judgment about particular cases with deduction from empirically validated laws (pp. 51–52). It will be clear that my account resolves each of these difficulties. Grünbaum draws an important distinction between (1) an asymmetry between explanation and prediction with regard to the grounds on which we claim to know that the explanandum is true, and (2) an asymmetry with respect to the logical relation between the explanans and the explanandum; he thinks that only the former sort of asymmetry obtains. I suggest that my account of the use of gappy laws will clarify both the sense in which Grünbaum is right (since an explanation and a tentative prediction can use similarly gappy laws which are similarly related to the known initial conditions and the result) and the sense in which, in such a case, we may contrast an entirely satisfactory explanation with a merely tentative prediction. Scriven (in his most recent statement, the review cited in n. 10 above) says that "we often pin down a factor as a cause by excluding other possible causes. Simple—but disastrous for the covering-law theory of explanation, because we can eliminate causes only for something *we know has occurred*. And if the grounds for our explanation of an event *have* to include knowledge of that event's occurrence, they cannot be used (without circularity) to predict the occurrence of that event" (p. 414). That is, the observation of this event in these circumstances may be a vital part of the evidence that justifies the particular causal explanation that we give of this event: it may itself go a long way toward establishing the elliptical law in relation to which we explain it (as I have shown in § 5), whereas a law used for prediction cannot thus rest on the observation of the event predicted. But as my account also shows, this does not introduce an asymmetry of Grünbaum's second sort, and is therefore not disastrous for the covering-law theory.

³⁷ See, for example, G. E. M. Anscombe, *Intention* (Oxford, 1957), especially p. 16; J. Teichmann, "Mental Cause and Effect," *Mind*, vol. 70 (1961), pp. 36–52. Teichmann speaks (p. 36) of "the difference between them and ordinary (or 'Humean') sequences of cause and effect" and says (p. 37) "it is sometimes in order for the person who blinks to say absolutely dogmatically that the cause is such-and-such, and to say this independently of his knowledge of any previously established correlations," and again "if the noise is a cause it seems to be one which is known to be such in a special way. It seems that while it is necessary for an observer to have knowledge of a previously established correlation between noises and Smith's jumpings, before he can assert that one causes the other, it is not necessary for Smith himself to have such knowledge."

II. SYMPOSIUM ON INDUCTIVE EVIDENCE

A. The Concept of Inductive Evidence

WESLEY C. SALMON*

I

HUME, it is often said,¹ tried to find a way of proving that inductive inferences with true premisses would have *true* conclusions. He properly failed to find any such justification precisely because it is the function of *deduction* to prove the truth of conclusions on the basis of true premisses. Induction has a different function. Given true premisses, the inductive inference establishes its conclusions as *probable*. Small wonder that Hume failed to find a justification of induction. He was trying to make induction into deduction, and he really succeeded only in proving the platitude that induction is not deduction.

Arguments along this line are appealing, and they have given rise to widely accepted attempts to dissolve the problem of induction. To understand and evaluate them we must, however, take account of ambiguities of the term "probable."

One important type of probability concept identifies probability with relative frequency. If we were to claim that inductive conclusions are probable in this sense, we would be claiming that inductive inferences with true premisses often have true conclusions, though perhaps not always. Unfortunately, Hume's argument shows directly that this claim cannot be substantiated. It was recognized long before Hume that inductive inferences cannot be expected *always* to lead to the truth. The suggestion that Hume merely showed the fallibility of induction is a mistake.² Hume's

argument shows not only that we cannot justify the claim that *every* inductive inference with true premisses will have a true conclusion, but further that we cannot prove that *any* inductive inference with true premisses will have a true conclusion. We can show neither that inductive inferences establish their conclusions as true, nor that they establish their conclusions as probable in the frequency sense. The introduction of the frequency concept of probability gives no help whatever in circumventing the problem of induction, but this is no surprise, for we should not have expected it to be suitable for this purpose.

A more promising probability concept identifies probability with degree of rational belief. To say that a statement is probable in this sense means that one would be rationally justified in believing it; degree of probability is the degree of assent a person would be rationally justified in giving. Probability is a logical relationship objectively determined by the available evidence. To say that a statement is probable in this sense means that it is supported by evidence. But, if a statement is the conclusion of an inductive inference it is supported by evidence—by inductive evidence—this is what it *means* in this context to be supported by evidence. Trivially, then, the conclusion of an inductive inference is probable under this concept of probability. To ask, with Hume, if we should accept inductive conclusions is tantamount to asking if we should fashion our beliefs in terms of the evidence, and this, in turn, is tantamount to asking

* This paper and the following comments were presented, with minor modifications, as a Symposium on Induction and Probability at the meetings of the American Philosophical Association, Western Division, Milwaukee, Wisconsin, April 30–May 2, 1964. The author wishes to express his gratitude to the National Science Foundation for its support of his research on probability and induction.

¹ The view outlined and criticized in this section is a composite drawn from various sources and is not to be attributed to any single author. Among those who subscribe to some such view I would include A. J. Ayer, *Language, Truth and Logic* (New York, Dover Publications, Inc., 1952); Paul Edwards, "Russell's Doubts about Induction," *Mind*, vol. 58 (1949), pp. 141–163; Asher Moore, "The Principle of Induction," *Journal of Philosophy*, vol. 49 (1952), pp. 741–758; Arthur Pap, *Elements of Analytic Philosophy* (New York, The Macmillan Co., 1949) and *An Introduction to the Philosophy of Science* (New York, The Free Press of Glencoe, 1962); P. F. Strawson, *Introduction to Logical Theory* (London, Methuen & Co. Ltd., 1952).

² See, for example, Jerrold Katz, *The Problem of Induction and Its Solution* (Chicago, University of Chicago Press, 1962), p. 115.

whether we should be rational. In this way we arrive at an "ordinary language dissolution" of the problem of induction. Once we understand clearly the meanings of such key terms as "rational," "probable," and "evidence" we see that the problem arose out of linguistic confusion and evaporates into the question of whether it is rational to be rational. Such tautological questions, if meaningful at all, demand affirmative answers.

Unfortunately, the dissolution is not satisfactory. Its inadequacy can be exhibited by focusing upon the concept of inductive evidence and seeing how it figures in the foregoing argument.³ The fundamental difficulty arises from the fact that the very notion of inductive evidence is determined by the rules of inductive inference. If a conclusion is to be supported by inductive evidence it is necessary that it be the conclusion of a correct inductive inference with true premisses. Whether the inductive inference is correct depends upon whether the rule governing that inference is correct. The relation of inductive evidential support is, therefore, inseparably bound to the correctness of rules of inductive inference. In order to be able to say whether a given statement is supported by inductive evidence we must be able to say which inductive rules are correct.

For example, suppose that a die has been thrown a large number of times and we have observed that the side two came up in one-sixth of the tosses. This is our "evidence" e . Let h be the conclusion that, "in the long run," side two will come up one-sixth of the times. Consider the following three rules:

- (1) (Induction by enumeration) Given m/n of observed A are B , to infer that the "long run" relative frequency of B among A is m/n .
- (2) (*A priori* rule) Regardless of observed frequencies, to infer that the "long run" relative frequency of B among A is $1/k$, where k is the number of possible outcomes—six in the case of the die.
- (3) (Counter-inductive rule) Given m/n of observed A are B , to infer that the "long run" relative frequency of B among A is $(n - m)/n$.

Under rule (1), e is positive evidence for h ; under rule (2), e is irrelevant to h ; and under rule (3), e is negative evidence for h . To determine which conclusions are supported by what evidence, it is necessary to arrive at a decision as to what inductive rules are acceptable. If rule (1) is correct, the evidence e supports the conclusion h . If rule (2) is correct, we are justified in drawing the conclusion h , but this is entirely independent of the observational evidence e ; the same conclusions would have been sanctioned by rule (2) regardless of observational evidence. If rule (3) is correct, we are not only prohibited from drawing the conclusion h , but also we are permitted to draw a conclusion h , which is logically incompatible with h . Whether a given conclusion is *supported by evidence*—whether it would be *rational to believe* it on the basis of given evidence—whether it is *made probable* by virtue of its relation to give evidence—depends upon selection of the correct rule or rules from among the infinitely many rules we might conceivably adopt.

The problem of induction can now be reformulated as a problem about evidence. What rules ought we to adopt to determine the nature of inductive evidence? What rules provide suitable concepts of inductive evidence? If we take the customary inductive rules to define the concept of inductive evidence, have we adopted a proper concept of evidence? Would the adoption of some alternative inductive rules provide a more suitable concept of evidence? These are genuine questions which need to be answered.

We find, moreover, that what appeared earlier as a pointless question now becomes significant and difficult. If we take the customary rules of inductive inference to provide a suitable definition of the relation of inductive evidential support, it makes considerable sense to ask whether it is rational to believe on the basis of evidence as thus defined rather than to believe on the basis of evidence as defined according to other rules. For instance, I believe that the *a priori* rule and the counter-inductive rule mentioned above are demonstrably unsatisfactory, and hence, they demonstrably fail to provide a suitable concept of inductive evidence.⁴ The important point is that

³ Wesley C. Salmon, "Should We Attempt to Justify Induction?" *Philosophical Studies*, vol. 8 (1957), pp. 33-48 contains a similar criticism of this type of argument. In that discussion I focused upon the concept of rationality; here the main emphasis is upon the concept of inductive evidence. Barker and Kyburg, in their comments on the present paper, bring the concept of rationality to the forefront again.

⁴ See Wesley C. Salmon, "Regular Rules of Induction," *Philosophical Review*, vol. 65 (1956), pp. 385-388, and "Inductive Inference" in *Philosophy of Science: The Delaware Seminar II*, ed. by Bernard H. Baumrin (New York, John Wiley & Sons, 1963).

something concerning the selection from among possible rules needs demonstration and is amenable to demonstration.

There is danger that we may be taken in by an easy equivocation. One meaning we may assign to the concept of inductive evidence is, roughly, the basis on which we ought to fashion our beliefs. Another meaning results from the relation of evidential support determined by whatever rule of inductive inference we adopt. It is only by supposing that these two concepts are the same that we suppose the problem of induction to have vanished. The problem of induction is still there; it is the problem of providing adequate grounds for the selection of inductive rules. We want the relation of evidential support determined by these rules to yield a concept of inductive evidence which is, in fact, the basis on which we ought to fashion our beliefs.

The foregoing problem is not circumvented by replacing rules of induction by confirmation functions. Confirmation functions determine evidential relations. Carnap's presentation of a continuum of inductive methods drives this point home. The problem is precisely the same. How are we to decide which confirmation function from among this superdenumerable infinity of confirmation functions provides a suitable concept of inductive evidence?

We began this initially promising approach to the problem of the justification of induction by introducing the notion of probability, but we end with a dilemma. If we take "probability" in the frequency sense, it is easy to see why it is advisable to accept probable conclusions in preference to improbable ones. In so doing we shall be right more often. Unfortunately, we cannot show that inferences conducted according to any particular rule establish conclusions that are probable in this sense. If we take "probability" in a nonfrequency sense it may be easy to show that inferences which conform to our accepted inductive rules establish their conclusions as probable. Unfortunately, we can find no reason to prefer conclusions which are probable in this sense to those which are im-

probable. As Hume has shown, we have no reason to suppose that probable conclusions will often be true and improbable ones will seldom be true. This dilemma is Hume's problem of induction all over again. We have been led to an interesting reformulation, but it is only a reformulation and not a solution.

II

In view of the multiplicity of possible inductive rules and the extreme difficulty in providing any sort of justification for the selection of one rather than another, it is tempting to hunt for still other ways of avoiding the task. One frequent theme begins by citing the ultimacy of the principles for which a justification is sought. It is impossible, we are told, to justify all principles, for in the absence of principles which can be used for purposes of carrying out the justification there is no conceivable means for achieving it. We are reminded that it is impossible to provide a justification of deduction, for, as we have learned from Lewis Carroll's "What the Tortoise said to Achilles,"⁵ the attempt will come to grief in a circle or an infinite regress. If we insist upon a justification for induction while we are content to omit the requirement for deduction, we are showing unseemly prejudice against one kind of inference. The fact that induction is not deduction does not mean that induction is inferior to deduction, or that induction stands in special need of justification.⁶

The first step in evaluating this kind of argument consists in becoming clear on the lesson to be learned from the colloquy between Achilles and the tortoise. This lesson, very simply, is that no supply of tautologies or logical truths can sanction the drawing of a conclusion from premisses unless there is a rule of inference whose function is to sanction inferences. *Modus ponens*, for instance, is a rule; the fact that there is a corresponding tautology in the propositional calculus does not make the rule dispensable. If there is any problem of justification here, it is the problem of justifying the adoption of a rule, which is neither true nor

⁵ *The Complete Works of Lewis Carroll* (New York, The Modern Library, 1936), pp. 1225 ff.

⁶ Again, the argument sketched is a composite. Among those who contribute to this standpoint are A. J. Ayer, *The Problem of Knowledge* (Penguin Books, 1956); Stephen Barker, *Induction and Hypothesis* (Ithaca, Cornell University Press, 1957); Max Black, *Language and Philosophy* (Ithaca, Cornell University Press, 1949), and *Problems of Analysis* (Ithaca, Cornell University Press, 1954); Rudolf Carnap, "The Aim of Inductive Logic," *Logic, Methodology and Philosophy of Science*, ed. by Ernest Nagel, Patrick Suppes, and Alfred Tarski (Stanford, Stanford University Press, 1952), and "Replicas and Systematic Expositions," § 26, *The Philosophy of Rudolf Carnap*, ed. by Paul Arthur Schilpp (LaSalle, The Open Court Publishing Co., 1953); Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, Mass., Harvard University Press, 1955); Katz, *op. cit.*; Henry E. Kyburg, Jr., *Probability and the Logic of Rational Belief* (Middletown, Conn., Wesleyan University Press, 1961); Strawson, *op. cit.*

false, rather than a problem of establishing the truth of a statement. If Achilles had had the wit to point out that his need was not for additional premisses, but rather for a rule of inference, the conversation might have taken a different and more constructive turn. The tortoise, it might be noted, was not reticent about accepting the pronouncements of logic; he was not questioning the legitimacy of accepting certain statements as *truths*.

The sole ground, it seems to me, for accepting or rejecting rules is in terms of the aims that will or will not be achieved by conforming to them. The aim of deductive logic, I take it, is to be able to draw true conclusions from true premisses. To achieve this aim we endeavor to adopt as rules of deductive inference only those rules which are truth preserving. We accept *modus ponens* as a rule of deductive inference because we believe it will never sanction drawing a false conclusion from true premisses. But can we justify *modus ponens* as a rule of deductive logic by proving that it is truth preserving? We can, of course, prove a metatheorem to the effect that *modus ponens* in the object language is truth preserving. The metaproof, however, requires inference in the metalanguage, and this in turn requires that the metalanguage have *modus ponens* or some other rule of inference which is at least as suspect. To prove that a kind of rule is truth preserving, it is necessary to have and use rules of deduction in that very proof. Hence, we cannot prove, without either circularity or vicious regress, that *modus ponens* is truth preserving—that is to say, we cannot justify deduction.

In view of this situation, it is sometimes said that we must utilize a kind of deductive intuition or logical common sense to "see" that *modus ponens* is a correct and legitimate form of inference. Then, the argument continues, if we are willing to allow the necessity of deductive intuition and common sense as the ultimate appeals in justifying deductive logic, why should we not allow a similar role to inductive intuition and inductive common sense? If an individual is so deductively blind that he cannot see the legitimacy of concluding *q* from *p* and if *p* then *q*, then there is nothing we can do with or for him. He simply does not belong to the community of sane and rational individuals. Communication, in any form, with such a person is impossible. The task of deductive logic is not to try to justify such rules as *modus ponens* but rather to present systematically the acceptable rules of deductive inference.

Similarly, the argument continues, if a person is

so inductively blind that he cannot see the superiority of induction by enumeration, i.e., rule (1), over the *a priori* rule (2) and the counter-inductive rule (3) there is nothing to be done. We can no more logically compel a person lacking in inductive common sense to accept acceptable rules than we can in the case of deduction. The task of inductive philosophy is not the justification of induction, but the formulation in systematic fashion of our inductive intuitions. This formulation may be no easy task, incidentally, for our intuitions may lie deeply hidden and may be in apparent conflict with one another, but sufficient critical reflection should enable us to straighten them out. In any case, the ultimate appeal for the justification of inductive rules is our intuitive sense for the concept of inductive evidence.

Persuasive as the foregoing analogy may seem, I think it is fundamentally misleading and unsound. The reason is simple. In the case of *modus ponens* I may reflect upon the rule very carefully and think of all sorts of instances of its application. Try as I will, I cannot conceive the possibility of any situation in which its use would lead from true premisses to a false conclusion. So deeply ingrained is this conviction that I am inclined to declare that any alleged counter example to *modus ponens* would *ipso facto* have to involve an equivocation or a misuse of the conditional form of statement. Part of what we mean by a conditional statement is that it shall be the valid major premiss of *modus ponens*. However, there is no need to dogmatize in advance. We cannot prove, without circularity, that no counter-example to *modus ponens* will ever be found. If an apparent counter-example is discovered it will receive the most careful scrutiny. Although we cannot now conceive the possibility of a genuine counter-example, future developments might make us change our minds. For the present, however, we can find no grounds whatever for withholding the judgment that *modus ponens* is truth preserving.

The situation with induction is quite different. After a small dose of Hume's *Enquiry* we can, without difficulty, imagine all sorts of states of affairs in which practically all—if not absolutely all—of our future inductive inferences with true premisses turn out to have false conclusions. We can, furthermore, construct perverse kinds of inductive rules (as judged in the light of our inductive intuitions) and describe possible worlds in which these rules would be very successful indeed. We cannot provide, without circularity, any reason

for supposing that we do not, in fact, live in some such world. A Cartesian demon could addle our brains and throw us into linguistic confusion, but we cannot conceive of anything he could do to make *modus ponens* not literally truth preserving. He would, however, have no trouble in completely subverting any inductive rule we can set forth. It is one thing to appeal to logical intuition concerning the acceptability of rules which, to the best of our most critical reflection, must be truth preserving in all possible worlds. It is quite another to make such an appeal to our (inductive) logical intuition for the acceptability of inductive rules when critical reflection reveals that they may turn out to be entirely unsatisfactory in our actual world and distinctly less successful than others we can formulate. There are live options in the case of inductive rules quite unlike any which exist in the case of rules of deduction.⁷

There is a close connection between the argument just discussed and that taken up in the first section of this essay. Our usage of terms like "rational," "probable," and "evidence" is closely linked to our inductive intuitions and our common sense of the inductive evidential relation. All were learned at mother's knee, and all are second nature to us. We can, nevertheless, state alternative and conflicting inductive rules, we can describe different linguistic usage, and we can imagine radically different inductive behavior. The problem is: Can we give adequate reasons for preferring our usage, intuitions, and behavior to the alternatives? If we can, that is a justification of induction. If we cannot, then we cannot justify induction. It is not, however, an adequate answer to say that we prefer them just because they are ours. Yet it seems to me that this is precisely what we are saying if we say that the generally accepted inductive rules stand in no need of justification, or if we say that our inductive intuition and our ordinary usage are the ultimate standards in terms of which we must justify. To say these things is to rule out the alternatives without a hearing, and to reject them simply because they fail to conform to what we already accept. Such an approach bespeaks a regrettable unwillingness to apply critical standards to cherished principles.

It may be objected that I am forgetting the ultimacy of the principles or rules under discussion, and the consequent impossibility of furnishing any kind of justification for them. In making precisely this point, Ayer remarks, "When it is understood that there logically could be no court of superior jurisdiction, it hardly seems troubling that inductive reasoning should be left, as it were, to act as judge in its own cause."⁸ The difficulty is, to pursue the metaphor, that there are too many courts of equal jurisdiction. Vicious circularity manifests itself; if each type of reasoning is left "to act as judge in its own cause," there are many conflicting judgments. For instance, in the court of affirming the consequent, the following argument would seem to lead conclusively to a verdict:

If affirming the consequent is valid,
then $2 + 2 = 4$.

$2 + 2 = 4$.

∴ Affirming the consequent is valid.

This argument has true premisses, it conforms to the form of affirming the consequent, and it asserts the validity of this form.⁹ If we allow customary inductive methods to act as judge in their own behalf, we do so by ignoring all other judges and listening only to the judgment in the particular court we happen to be in.

Strawson also has seen fit to make a legal reference in dealing with the same point: "But it makes no sense to inquire in general whether the law of the land, the legal system as a whole, is or is not legal. For to what legal standards are we appealing?"¹⁰ This analogy is a useful one. It is, indeed, pointless to ask whether the legal system as a whole is legal, but this does not mean that the legal system is exempt from criticism and stands in no need of justification. What does make sense is to ask whether adherence to this legal system will achieve ends we seek to realize, and whether some other legal system would achieve these ends more efficiently. Similarly, although we cannot justify inductive rules by reference to other inductive rules, we can try to show that there are reasons for preferring one inductive rule to others.

⁷ The choice between a two-valued and a three-valued logic is not at all the same sort of thing. Two-valued logic is a specialization of three-valued logic. No such relation exists in the case of inductive rules; here there is genuine conflict. The choice between *modus ponens* and affirming the consequent would be more to the point.

⁸ Ayer, *The Problem of Knowledge*, *op. cit.*, p. 75.

⁹ Salmon, "Should We Attempt to Justify Induction," *op. cit.*, presents inductive examples.

¹⁰ Strawson, *op. cit.*, p. 257.

II. SYMPOSIUM ON INDUCTIVE EVIDENCE

B. Discussion: Is There a Problem of Induction?

STEPHEN F. BARKER

WITTGENSTEIN, Strawson, and others have held that the traditional problem of induction is a pseudo-problem, resulting from conceptual confusion; a puzzle to be dissolved, not a problem to be solved in its own terms. Professor Salmon disagrees and tries to rescue the grand old problem from dissolution; or perhaps I ought rather to say that he tries to resurrect that grand old corpse of a problem which many of us had hoped would now be allowed to molder in peace.

What is this problem of induction, as Salmon sees it? It involves a contrast between deduction and induction. As I understand him, Salmon admits that our practice of reasoning deductively—our practice of preferring *modus ponens* to affirming the consequent, and so on—is a practice which, considered as a whole, is something we cannot justify. Salmon thinks that any attempt at justification would itself employ deductive reasoning, and therefore would fail through begging the question. Nevertheless, Salmon does not feel that this makes deduction an irrational practice. He does not doubt that the practice of deductive reasoning is suitable, satisfactory, proper, adequately grounded, and a practice we ought to adhere to. For it seems to him that when we reflect upon any one of our specific rules of deduction—for instance, *modus ponens*—we find it is inconceivable that the rule could lead us from true premisses to a false conclusion.

What sort of inconceivability is involved here? Salmon expresses himself cryptically. But surely the sense in which it is inconceivable that *modus ponens* could be unreliable is not like the sense in which it is inconceivable that a man now living could be 100 billion years old. People with more flexible and more versatile imaginations than ours might well be able to conceive of a man's being 100 billion years old, but even they could not conceive of *modus ponens* as being unreliable. No, the sense in which it is inconceivable that *modus ponens* could be unreliable must be like the sense in which it is inconceivable that there could be

spherical pyramids. With shapes, being a sphere necessarily precludes being a pyramid, and with arguments being in the form *modus ponens* necessarily precludes being unreliable. Reflection upon the senses we attach to these terms enables us to see that this is necessarily so.

But now for induction. Our practice of reasoning inductively can very crudely be described as the practice of attributing to unobserved phenomena the simplest regularities that we have detected in observed phenomena; or, still more crudely, it can be described as the practice of obeying the so-called rule of induction by simple enumeration. With regard to this practice, Salmon feels qualms not felt about deduction. He maintains that there is a crucial difference between the status of induction and that of deduction. Deduction is safe, since it is inconceivable that it could go wrong; but induction may be unsafe, according to Salmon, for we can conceive of its going wrong.

What would it be like for the practice of induction to go wrong? Well, now, we can conceive of people who made it their practice to reason in some anti-inductive manner—they always attribute to the unobserved something other than the simplest regularity detectable in the observed. To suppose that induction could go wrong, according to Salmon, is to suppose that the long-run relative frequency with which people who follow the practice of reasoning inductively get true conclusions from true premisses is less than the long-run relative frequency with which people who follow the practice of reasoning in some anti-inductive manner do so. Perhaps the world is under the governance of a malicious Cartesian demon who makes it his sport to see to it that people who reason inductively will usually reach false conclusions even when their premisses are true, and who sees to it that people who reason in some anti-inductive manner usually get true conclusions when their premisses are true. In such a world induction would not pay. Now, how do we know that our actual world is not like this? We can

conceive that induction might not be effective as a means of leading us to truth. Thus induction may be unsafe in a way in which deduction cannot be. This is what leads Salmon to believe that we need a justification of induction, a justification of a sort not called for in connection with deduction. We need some account of why it is rational to adopt the practice of inductive reasoning rather than some anti-inductive practice. We need some grounds for believing that induction is a suitable, satisfactory, proper practice, adequately grounded, that we ought to pursue. (These words "rational," "suitable," "satisfactory," etc., are the terms Salmon uses.)

So far, I have been summarizing what I think Salmon is saying. Now I have several criticisms. My first and main objection to Salmon's way of trying to build up a problem here is that he neglects or underrates a most important certainty that we do possess regarding induction. I fully agree with him that it is conceivable that induction might be less successful than some other way of reasoning about the world; this is conceivable, and it is a logical possibility. But it is not probable. We know for an absolute certainty that it is not probable that any anti-inductive practice will be as successful in the long run as induction will be. Here of course I am using the term "probable" in the sense of rational credibility, not in the relative frequency sense. Built into this normal sense of the word "probable" is a commitment to the practice of induction. To be sure, we do not know with certainty that people who practice induction will be more successful in reaching true conclusions than will those who practice some form of anti-induction; but what we do know with certainty is that those who practice induction will probably be more successful—that is, that it is reasonable to believe that they will be more successful. That this is so reflects an aspect of what the word "probable" means in its normal sense. Just as it is inconceivable that *modus ponens* should be unreliable, so it is inconceivable that inductive inferences should not *probably* be the most successful kind in the long run. But to grant this is to grant that we do have good reason for regarding induction as reliable. This means that there is no problem of induction, in Salmon's sense. There is no more of an over-all problem about the reliability of induction than there is an over-all problem about the reliability of deduction.

Salmon will reject what I have just said, for he thinks of it as a shuffling evasion. He will reply

that even if induction is probably reliable, even if this is necessarily true in virtue of the normal meaning of the word "probable," still there remains a serious problem of induction. The problem is: Why is the fact that induction is probably reliable any real reason for relying upon induction? Why should we prefer probable conclusions to improbable ones?

But I am afraid this won't do. The question "Why should we believe probable conclusions?" still is not a successful way of posing a problem of induction. For the question "Why believe probable conclusions?" is too much like the questions "Why is the beauty of a thing a reason for admiring it?" "Why is the immorality of an action a reason for refraining from it?" and "Why is the goodness of a thing a reason for desiring it?" All these questions, when asked earnestly and tenaciously, arise from conceptual confusions. It is a mistake to search earnestly for some deep proof that the beauty of a thing constitutes a real reason for admiring it; this is a mistake because when I say that a thing is beautiful I am not merely describing it: also an essential part of what I am doing is taking a stand in favor of admiring it. It is mistaken to search earnestly for a proof that immoral actions are to be avoided, since to say that an action is immoral is not merely to describe it: also an essential part of what one is doing is taking a stand against performing it. It is mistaken to ask earnestly for proof that the good is to be desired, since in saying that a thing is good I am not merely describing it: also an essential part of what I am doing is taking my stand in favor of desiring it. Similarly, it is mistaken to ask earnestly whether probable conclusions are to be preferred, since in saying that a conclusion is probable one is not merely describing it: also an essential part of what one is doing is taking one's stand in favor of believing it. Thus it is inconsistent (in a broad sense) to deny that probable conclusions are to be preferred.

Salmon advocates a radical questioning of induction; he wants to question whether the whole practice of induction is not perhaps mistaken, root and branch. What I am arguing is that he has not succeeded in framing a coherent question. Moreover, I am suggesting that there is no coherent radical question needing to be asked here, for the general practice of induction is justified simply by the fact that its conclusions are probable, and it is inconsistent to deny the preferability of probable beliefs. Of course this does not mean that we are necessarily bound blindly to accept all conventional

inductive practices, any more than we are bound to accept all conventional moral practices. Of course it can be coherent to question and to criticize particular inductive procedures, just as it can be coherent to question and to criticize particular prevailing practices. But we do this by contrasting the given practice with other accepted practices, trying to show that it does not gibe with them. Such internal criticism of induction aims at a harmonious reconciliation of conflicting tendencies that may be present in our inductive practices. But this is very unlike the radical external criticism of induction advocated by Salmon.

As Salmon frames his problem, it is the question whether induction is more "rational," "suitable," "satisfactory," and "proper" than are anti-inductive methods. It is important to notice that these words are inadequate to carry the load Salmon imposes upon them. Already built into the normal sense of the word "rational" is a reference to inductive standards; in the normal sense of the word "rational," a rational man is necessarily one who, among other things, reasons inductively rather than anti-inductively. The man who expects for the future the opposite of what he has observed in the past is a paradigm case of irrationality: we point to him when we teach children the meaning of the term "irrational." The words "proper," "suitable," and "satisfactory" are no better for Salmon's purposes, for what can they mean in this context except rational, or probably reliable? The point is that the senses of this whole family of words are permeated by commitment to the

practice of induction, a practice which shapes our entire form of life. Salmon wants to call into question the legitimacy of this form of life; yet such a question cannot be coherently formulated in our language, or in any language, spoken by persons like ourselves. Salmon would like us radically to question the practice of induction which shapes our whole form of life, but words fail. We cannot express such a question. We reach one of those points at which, as Wittgenstein says, one feels like uttering an inarticulate cry. And I say that the question which cannot be put, far from revealing the essence of man (as the darker philosophers would have it), actually reveals itself not to be a question.

One final comment. I think Professor Salmon tends to be slightly unfair when he in several places intimates that to regard the problem of induction as deserving dissolution rather than solution is to regard it as trivial. That is not so. One can regard the problem of induction as a conceptual confusion and yet still regard it as a deep and important confusion. There is nothing shallow or trivial about the problem as it appears in Hume's thought, and it is greatly to Hume's credit that he had the intellectual penetration without which he could not have fallen into his conceptual difficulties about induction. We do not necessarily denigrate a philosopher's achievement when we say that he was a victim of conceptual confusion. And we do not necessarily waste our own time when we devote lengthy study to the unraveling of pseudo-problems.

The Ohio State University

II. SYMPOSIUM ON INDUCTIVE EVIDENCE

C. Comments on Salmon's "Inductive Evidence"

HENRY E. KYBURG, JR.

I WOULD like to begin by saying that I think Salmon's reading of Hume's sceptical result is impeccable. It is indeed so clear and so faultless, that I would like to repeat it to you: "Hume's argument shows not only that we cannot justify the claim that *every* inductive reference with true premisses will have a true conclusion, but further that we cannot prove that *any* inductive inference with true premisses will have a true conclusion." This statement is clearly an important one to keep in mind when we talk about induction, inductive evidence, the justification of induction, inductive logic, and such things. Although I agree in the main with much of what Salmon says (indeed, with very nearly everything in the first section of his paper), there are places where I think he strays from the straight and narrow, and it will turn out that those places are precisely those places where he has inadvertently forgotten the conclusion of Hume's argument.

The problem of induction that Salmon is willing to take seriously—the problem which we may consider the new and valid problem of induction—is the problem of "providing adequate grounds for the selection of inductive rules." It is indeed the case, as Salmon himself has shown with elegance and finality, that some proposed inductive rules are better than others—that, for example, the counter-inductive rule is, in a sense, unsatisfactory. (It is unsatisfactory because it leads to conflicting predictions.) It is perfectly true that something concerning the selection from among possible rules needs demonstration and is amenable to demonstration. It is also true that Carnap has defined a whole continuum of inductive methods, and that the confirmation functions discussed there do not settle the problem of choosing an inductive rule.

On what basis, then, are we to choose among inductive rules? Salmon proposes that there is a basis on which to choose, and a standard according to which to choose; and that to provide this standard (which of course will lead to the conventional rule which we all follow anyway) is a

solution to "Hume's problem of induction." The basis on which we choose is the deductively established character of each of the possible inductive rules, and the standard is to be pragmatic. This kind of justification of induction, he calls (following Black) the "practicalist's" justification.

A good part of Salmon's paper concerns a supposed analogy between the justification of the ultimate *deductive* rules, and the justification of the ultimate *inductive* rules. It has been maintained, on the basis of this analogy, that ultimate inductive rules can only be justified on an intuitive basis, just as (it is claimed) ultimate deductive rules can only be justified on an intuitive basis. Salmon attacks that analogy; I shall attempt to defend it.

Feigl, long ago, introduced an appetizing distinction between validation and vindication. We can validate some deductive rules by reference to more "ultimate" deductive rules; presumably we can validate (inductively) certain inductive rules by reference to more "ultimate" inductive rules. But when it comes to the ultimate rules themselves, we can no longer seek *validation* (this would be self-contradictory, for validation *means* "by reference to more ultimate rules"); we must seek *vindication* instead: that is, we must seek arguments to show that these ultimate rules will accomplish the purposes they are supposed to accomplish.

Salmon says (following Feigl) that the aim of deductive rules is to be able to draw true conclusions from true premisses. This is one desideratum. Even better, of course, would be to have deductive rules that enable us to draw true conclusions from *any* premisses. Furthermore, there *are* such deductive rules, e.g., "From P , infer $2 = 2$." But of course this desideratum, it is easy to show anyone except an absolute idealist or a dialectical materialist, is *inappropriate*, in the sense that we can't construct a system of deductive logic satisfying it which will also do the other things we want it to do. For example, we want it to organize our discourse in certain ways; we want it to be

powerful enough to lead to the validation of our customary deductive rules of inference (the ones we learned at mother's knee).

One ultimate deductive rule that does seem to satisfy a large number of desiderata is *modus ponens*. (Of course it is only "ultimate" in a particular system of logic—other systems might take other rules as "ultimate"—but this leads to problems that I am willing to bypass.) The one desideratum that Salmon mentions for deductive rules is that they be truth preserving. And he admits that we cannot *prove* that *modus ponens* is truth preserving. Although we cannot prove that it is truth preserving, we can reflect on it very carefully, and observe that it is difficult, impossible even, to "conceive the possibility of any situation in which [its] use would lead from true premisses to false conclusions." These are Salmon's words.

But, he says, the situation with regard to inductive rules is quite different. I quote: "After a small dose of Hume's *Enquiry* we can, without difficulty, imagine all sorts of states of affairs in which practically all—if not absolutely all—of our future inductive inferences with true premisses turn out to have false conclusions." We are, then, applying precisely the same standard to inductive rules that we applied to deductive rules: they are to be truth preserving. But even the little dose of the *Enquiry* that appears early in Salmon's paper suffices to convince us that this standard is altogether *inappropriate* to inductive inference. It is inappropriate in precisely the way that the demand that deductive inference yield only true conclusions is inappropriate. Of course what we really want from deductive inference is the truth. Why hedge? Because, as we all know perfectly well, deductive rules with this highly desirable property, which are also deductively powerful and interesting, simply *are not available*. In precisely the same way, Hume's argument (the very one mentioned by Salmon) shows that inductive rules with the weaker but still desirable property (which does happen to be appropriate to deductive rules) that they lead from true premisses only to true conclusions, or even inductive rules that we know have the property that they *sometimes* lead to true conclusions from true premisses, *simply are not available*. They are no more available than acceptable deductive rules that are truth insuring. And there's an end of the matter.

Since this is an inappropriate standard, and since Salmon, Hume, and I all agree that it is an inappropriate standard, it is clearly utterly ir-

relevant for Salmon to cloud the issue with the ghosts of Cartesian demons. *Of course* it is possible that inductive rules are not truth preserving. The analogy that Salmon destroys here is not the relevant one. The relevant analogy, in bald terms, is this: An ultimate deductive rule (*modus ponens*, for example), which can no longer be validated, since it is ultimate, can be vindicated by (and this is Salmon's word) *reflection*: by reflecting on the rule very carefully, and failing, nevertheless, to "conceive the possibility of any situation" in which it would fail to accomplish *what it may appropriately be expected to accomplish*. Similarly, an inductive rule which can no longer be validated, since it is ultimate (actually, here we should talk about the definition of probability that leads to the inductive rules) can be vindicated by reflecting on the rule very carefully, and failing, nevertheless, to conceive the possibility of any situation in which it would fail to accomplish that it may *appropriately* be expected to accomplish. Of course it can't appropriately be expected to be truth preserving. We can appropriately expect it to lead to self-consistent predictions; we can appropriately expect it to be powerful enough to yield interesting results, and so on. And this, I think, is a requirement that is perfectly well satisfied, for example, by my own definition of probability, barring a few wrinkles. It is certainly a requirement that was (originally) intended to be satisfied by Carnap's *c** (although things didn't work out that way). The appropriate thing to ask about an inductive rule (or a definition of probability) is not whether there is a universe where the inductive rule leads from true premisses to false conclusions (or where the definition of probability leads to inductive rules that are not truth-preserving), but whether we can conceive of a universe in which (for example) (1) all the *A*'s that an individual has seen have been *B*'s, (2) there is absolutely nothing else that that individual in that universe knows, and yet (3) it would be *irrational* for him to expect the next *A* to be a *B*. And this is indeed just the kind of argument that people who have proposed inductive notions of probability have used as litmus to test the soundness of their proposals. (The reference to the "universe" in this example is gratuitous, of course; and that is precisely the point.)

Now what I have just gone through does no more than to rehabilitate the analogy that Salmon claims to have destroyed. I think that the only reason he makes the analogy seem to fail is that (a) he considers only one desideratum that a rule

may satisfy, and (b) this desideratum is appropriate for deductive rules and inappropriate for inductive rules.

Salmon also says that the situation with respect to inductive rules is different from that with respect to deductive rules, because there are no live options for deductive rules. This isn't so. There is, e.g., the deductive rule I mentioned earlier; there is the deductive rule of affirming the consequent which Salmon mentions toward the end of his paper. There are no live options among deductive rules that do what we want them to do, and what we can legitimately expect them to do. True. And I don't think there are any live options among inductive rules (that is, among definitions of inductive probability) that do what we want them to do and can legitimately expect them to do.

Salmon asks if we can give reasons for preferring our conventional usage, intuitions, and behavior to the alternatives. This applies with equal force to deductive rules and to inductive rules. To a very large extent, we *can* give adequate reasons, and, as Salmon says, this is a justification of induction. I agree with him wholeheartedly in his general conclusion—with the important exception that I think that in some sense our justification of inductive rules must rest on an ineradicable element of inductive intuition—just as I would say our justification of deductive rules must ultimately rest, in part, on an element of deductive intuition: we *see* that *modus ponens* is truth preserving—this is simply the same as to reflect on it and fail to see how it can lead us astray. In the same way, we *see* that if all we know about in all the world is that all the *A*'s we've seen have been

B's, it is *rational* to *expect* that the next *A* will be a *B*.

There is one more loose end. Where does this leave the practicalist's justification of induction? I said that there were a number of desiderata that apply to inductive rules, just as there are a number of desiderata that apply to deductive rules. And I think that the practicalist's arguments serve very well to help select inductive rules (or theories of probability that lead to inductive rules) that meet these desiderata. Thus we may quite naturally want any rule of inductive estimation to have asymptotic properties; we may quite naturally desire that inductive rules lead to estimates that are consistent with each other (as the counter-inductive methods clearly do not).

It should be pointed out, of course, that these are *intuitive requirements*; they have *nothing at all* to do with our success in using these rules. Our use of the rules is necessarily only finite (in the long run we are all dead, as the first great inductive probabilist said), and in any finite run, we may, in point of fact, have better results with rules that are not asymptotic, and even with rules that actually lead to *inconsistent* estimates. What really counts, ultimately, even here on the practicalist's home ground, it strikes me, is inductive intuition. Our object should not be to try to replace that intuition with something else—in particular it should not be to try to replace it with just precisely *that* which Hume's argument tells us we can't replace it with—but it should be to attempt to reduce all the inductive rules that we use—those rules learned at our mother's knees—to the fewest and clearest and intuitively most acceptable principles we can possibly reduce them to.

Wayne State University

II. SYMPOSIUM ON INDUCTIVE EVIDENCE

D. Rejoinder to Barker and Kyburg

WESLEY C. SALMON

PROFESSORS Barker and Kyburg agree, though on different grounds, that induction is justified because it is rational. Barker argues for an analytic relation, via probability, between induction and rationality, while Kyburg claims that the connection between induction and rationality is intuitively clear. In sharp contrast to my discussants, I maintain that the concept of rationality will not do the job by itself, but that it needs help from another quarter. A justification of induction must, I hold, hinge upon a relation between induction and frequency of truth-preservation or success. I do not deny, of course, that induction is rational; I claim it is rational *because of* its relation to truth-preservation. Barker and Kyburg, on the other hand, assert that the use of inductive methods is rational *regardless of* any relation to success. Both agree that, no matter how infrequently inductive methods might yield true conclusions from true premisses, and no matter how frequently some perverse anti-inductive method might be successful, induction would still be justified because it is rational. I reject this answer.

Much of the burden of the controversy rests upon the analysis of the concepts of rationality and probability. It is perfectly true, as Barker emphasizes, that terms like "probable" and "rational" sometimes function as terms of cognitive appraisal. They are used to commend beliefs, assertions, propositions, etc. We are taught various criteria for the application of these terms by our parents, our teachers, and our society. The same holds for "good," "beautiful," "immoral," and other terms of ethical or aesthetic evaluation. To say that a painting is beautiful is (at least in part) to take a stand in favor of admiring it. To say that an act is immoral is (at least in part) to express disapproval of it. To say, in some contexts, that a conclusion is probable is (at least in part) to recommend its acceptance. Thus, our social group evolves a common morality, a common aesthetic, and a common methodology for appraisal

of factual beliefs. When an object satisfies certain descriptive characteristics we commend it aesthetically; when an act satisfies certain descriptive characteristics we commend it morally; when a conclusion satisfies certain descriptive (logical) characteristics we commend it cognitively. Furthermore, the usual canons of induction certainly do play a basic part in the criteria of cognitive evaluation.

This situation obviously gives rise to fundamental problems. We all recognize the importance of the moral (or aesthetic) critic. According to the common morality, a certain practice is acceptable, but the moral critic asks whether it is really acceptable. Cannibalism, murder, rape, and slavery have each been accepted, and even commended, by some prevailing moral code. We are more civilized than we would otherwise have been because men challenged, criticized, fought, and rejected the common morality. I cannot accept the view that the common morality is above criticism because of our ordinary use of words. For precisely the same reasons, I cannot agree that induction needs no justification just because it happens to conform to the common methodology.

There is a familiar ambiguity in terms like "good." One meaning stems from the common morality and merely reflects the accepted standards. To say that something is good in this sense means merely that it is commended by society in general. A morally deeper meaning is that of being *worthy of commendation* regardless of what society as a whole may think. When an act is good in the former sense, the question remains whether it is good in the latter sense. Does the fact that most people in a particular group call something "good" mean that it is good? Is an act (such as lynching or church bombing) which is commended in a given society actually deserving of commendation? Such questions had better not turn out to be meaningless!

Barker states explicitly that he does not embrace a view that rules out criticism of common moral

practices. He suggests, however, that such criticism must be piecemeal; it amounts to seeking some sort of coherence within the accepted morality. He expresses serious doubt about the possibility of calling the whole moral framework into question. This account seems to me inadequate. Consistency—even in a fairly extended sense—is not a sufficient basis for moral criticism; internal inconsistency alone cannot provide grounds for extensive moral reform. Despite many serious attempts to show otherwise, it seems entirely possible for an individual or a society to be highly immoral without violating any canons of logic. Furthermore, we can and do reject common practices, not on the basis of inconsistency, but on such grounds as simple abhorrence. Coherence theories of morality are, I suggest, just as inadequate as coherence theories of truth, and for just about the same reasons.

This obviously is not the place to discuss the grounds of moral judgment, so I shall say no more about the various meanings of moral terms. The fundamental ambiguity is clear enough. Barker has drawn an analogy between moral (or aesthetic) evaluation and cognitive evaluation. He regards this analogy as damaging to my view that induction stands in need of justification. I have tried to show that this analogy, like Strawson's legal analogy, tends to support the view that there is a real problem of justifying induction which is not amenable to dissolution by tracing the ordinary uses of words. We *can* ask whether moral practices commended by our society are worthy of commendation, and we *can* ask whether the inductive practices commended by our society are worthy of commendation.

The aesthetic, ethical, and legal analogies we have been discussing may have considerable persuasive value, but they can hardly have much probative force. The serious grounds for denying that there is a problem of justifying induction, while illustrated by these analogies, are quite distinct. Barker takes up two closely related ones. The first is that there are no more fundamental principles in terms of which the justification could be carried out, and the second is that we simply do not have the language in which to pose the problem coherently. Both of these arguments seem to me to suffer the fatal defect of failure to take

account of Feigl's extremely fruitful distinction between validation and vindication.¹ Feigl has shown that there is a kind of justification—vindication—which does not rely upon the existence of more fundamental justifying principles. Barker, however, seems to be maintaining that validation is the only kind of justification possible in ethics, aesthetics, or logic. He seems to be denying the very possibility of vindication, but in so doing he, like every other proponent of this argument I know of, simply ignores Feigl's important distinction. The only reason Barker seems to give for denying the possibility of vindication is that our language—which is, of course, an expression of our common sense, our common morality, and our common methodology—contains the ultimate validating principles. These principles are incapable of being vindicated because, apparently, they are so deeply ingrained in our language that we cannot coherently ask questions about their vindication.

While I agree with Barker that many of the words we want to use to discuss the problem of induction are "permeated by commitment to the practice of induction, a practice which shapes our entire form of life," it still seems to me that he underrates the power of our language. The language to which Barker refers is the language in which Hume wrote; it is the language in which Feigl stated the distinction between validation and vindication and showed how this distinction bears upon the problem of induction; it is the language in which we explore the ambiguities of a host of words such as "good," "beautiful," "immoral," "probable," and "rational"; it is the language in which Carnap presented a continuum of inductive methods, and in which I have on numerous occasions discussed alternative inductive rules. These are not "inarticulate cries!"

I am not saying that it is completely apparent what would constitute a vindication of induction. If we could prove that induction is frequently truth-preserving, that would provide an easy answer, but we all agree that no such proof is possible. Kyburg suggests, and Barker might agree, that we vindicate induction by showing that it leads to rational results even though it may not lead to true results. This answer is, unfortunately, difficult to make intelligible.

¹ Herbert Feigl, "De Principiis Non Disputandum . . ." in *Philosophical Analysis*, ed. by Max Black (Ithaca, Cornell University Press, 1950). In my discussions of the problem of induction I have made extensive use of Feigl's distinction; see, for example, "Should We Attempt to Justify Induction?" *Philosophical Studies*, vol. 8 (1957), pp. 33-48, and "Inductive Inference" in *Philosophy of Science: The Delaware Seminar*, vol. II, ed. by Bernard H. Baumrin (New York, John Wiley & Sons, 1963).

Consider some of the most usual ways in which the term "rational" and its cognates are used, noting particularly the situations in which we withhold the term. We deny that people are rational (and derivatively, their practices and their thinking) if they are exceedingly drunk, under the influence of certain drugs, just coming out of certain anesthetics, extremely young, psychotic, completely lacking in common sense and practicality, unaware of the most familiar matters of fact, or extremely deviant in belief or behavior. To a significant extent, "rational" connotes basic agreement with the user. Those who are politically far right are likely to regard those of the far left as irrational and vice versa, while the moderate is apt to doubt seriously the rationality of all extremists (except, perhaps, those who carry moderation to an extreme). We can tolerate some factual disagreements, but it is harder to tolerate methodological differences. There is hope of resolving factual disagreements if there is common methodological ground, but it is harder to resolve methodological differences. When methodological disagreements are really deep and irresolvable we are strongly tempted to challenge the rationality of the other: "You just can't reason with him!" To allow that a person is rational suggests that he does not differ methodologically to such an extent that reaching an agreement is insuperably difficult. Rationality involves a kind of social conformity.

The foregoing remarks about "rational" are more pragmatic than semantic. Literally, "rational" means "capable of reasoning correctly." People are called "rational" if we think they reason well; arguments are called "rational" if we think they are logically correct; and logical principles are called "rational" if we think they are canons of correct logical argument. We all agree that certain familiar inductive methods are logically correct, so we call them and the people who use them "rational." In this discussion, however, the very problem at issue is what constitutes a correct inductive method. The fundamental question arises for cognitive evaluation just as it did for moral evaluation: Are the things we actually commend really worthy of commendation? Are the inductive methods we regard as correct really correct?

I submit that we commend the principles of deductive logic because we believe they are invariably truth-preserving. I submit that we commend the accepted principles of inductive logic because we believe they are frequently truth-

preserving. We call such methods "rational" because we believe they are frequently truth-preserving. Hume shows us, of course, that we cannot prove that they are in fact frequently truth-preserving, so Barker and Kyburg conclude that frequency of truth-preservation cannot provide a basis for vindication. Instead, they claim, rationality is the basis for vindication. We can, I agree, show without difficulty that certain inductive methods are rational if this means only that they are methods which are thought to be frequently truth-preserving, but that is no vindication of an inductive method. Such socio-psychological facts have no bearing upon justification.

The question I am raising is how to construe "rational" in order to make sense of the thesis that induction is justified because of its rationality. Let me suggest several possibilities. (1) "Rational" might mean "logically correct" where "logically correct" means "frequently truth-preserving." With this meaning of "rational," as Barker and Kyburg rightly insist, inductive methods cannot be proved to be rational. (2) "Rational" might mean "regarded by most people in our circle as frequently truth-preserving." With this meaning of "rational," to show that an inductive method is rational is simply to establish the sociological fact that certain people think it is frequently truth-preserving. Unless we can find some ground for their opinion, this fact is without logical interest. (3) The meaning of "rational," it might be urged, is determined by the criteria for its application we have absorbed in learning to speak our language. This, I suspect, comes close to the sort of answer Barker would give. It does not differ fundamentally from (2), for the criteria we learn are just those which single out the methods which our parents, our teachers, and our society regard as frequently truth-preserving. Finally, (4) "rational" might mean "conforming to the principles commended by our group." This may or may not coincide with sense (2), for we may commend inductive methods because we believe them to be frequently truth-preserving or we may commend them for some other reason. Mere commendation by society is of no use for vindication unless we are more interested in social approval than in truth. If, however, there is some desideratum other than frequency of truth-preservation which underlies our commendation of inductive methods, then we want to bring it out and examine it. I do not know what it is. Barker and Kyburg have not told us what it is. If rationality is that

desideratum, then it is very important that we be told what rationality is. In what sense is the term "rational" now being used? "Conforming to accepted inductive methods" is no answer, for that would just bring us back to the view that accepted inductive methods have the desirable (?) characteristic of being accepted inductive methods.

Kyburg argues that we cannot dispense with inductive intuition. After carefully examining the logical properties of possible inductive methods, we may still have to rely on intuition much as we do in the deductive case. We see that certain inductive methods lead to *rational* expectations, and that is all there is to it. Reluctant as I am to admit inductive intuition, I must concede that it may be necessary in the last analysis to depend upon it. If inductive intuition does prove indispensable, then accept it we must, but we should make every effort to postpone its admission as far as possible and minimize its role. Furthermore, if inductive intuition is required, it will, I believe, have to be an intuition concerning a relation between induction and frequency of truth-preservation, not between induction and rationality as Kyburg would have it. I know what it means to intuit that a particular inductive method is frequently truth-preserving. Though such an intuition might be utterly unreliable, I could recognize it if I had it. If, however, the rationality of an inductive method had nothing whatever to do with frequency of truth-preservation, I really do not know what it would be like to intuit its rationality. Would it be simply a matter of intuiting that the use of the method would earn social approval, no matter how unsuccessful it might otherwise turn out? It is of first importance that those who claim that rationality is the aim of induction should tell us what rationality is.

It may seem that I have taken this discussion up a blind alley. While calling into question the possibility of vindicating induction on the basis of rationality alone, I have been insisting that the

only adequate ground for vindication is frequency of truth-preservation. We have all agreed from the outset, however, that we cannot prove that any inductive method is frequently truth-preserving. Where does this leave the argument? Kyburg has said that frequency of truth-preservation is not an appropriate desideratum for inductive methods, for methods which are ampliative and frequently truth-preserving are simply not available. This is not quite accurate. Methods which are ampliative and which can be demonstrated to be frequently truth-preserving are not available, but methods which are ampliative and frequently truth-preserving have been available, seem to be available now, and may be available in future. It is just that we cannot prove that any given one will turn out to have this very desirable characteristic. Nevertheless, we shall do our level best to find one that has it. The fact that we cannot prove that a given method will be frequently truth-preserving means only that the relationship between induction and frequency of truth-preservation by means of which induction is to be vindicated is more complex and indirect than we might have hoped before reading Hume.

Kyburg concludes his remarks by suggesting that the kinds of requirements I have elsewhere endorsed for vindicating induction have no connection with frequency of truth-preservation. He regards them as intuitive requirements of rationality. Again, I must disagree. Reichenbach's convergence requirement does have something to do with over-all success. To adopt a rule that violates the requirement is to insure endlessly recurrent error. Requirements of consistency also have to do with over-all success. To adopt an inconsistent inductive logic is to forego systematically any distinction between truth and error. To be sure, the recalcitrant problem of the short run remains unsolved, but that means only that the job of vindication is not yet done. This I gladly acknowledge, but I insist that the work is not superfluous.

III. OPERATIONALISM AND ORDINARY LANGUAGE A CRITIQUE OF WITTGENSTEIN

C. S. CHIHARA AND J. A. FODOR^{1,2}

INTRODUCTION

THIS paper explores some lines of argument in Wittgenstein's post-*Tractatus* writings in order to indicate the relations between Wittgenstein's philosophical psychology on the one hand and his philosophy of language, his epistemology, and his doctrines about the nature of philosophical analysis on the other. We shall hold that the later writings of Wittgenstein express a coherent doctrine in which an operationalistic analysis of confirmation and language supports a philosophical psychology of a type we shall call "logical behaviorism."

We shall also maintain that there are good grounds for rejecting the philosophical theory implicit in Wittgenstein's other works. In particular we shall first argue that Wittgenstein's position leads to some implausible conclusions concerning the nature of language and psychology; second, we shall maintain that the arguments Wittgenstein provides are inconclusive; and third, we shall try to sketch an alternative position which avoids many of the difficulties implicit in Wittgenstein's philosophy. In exposing and rejecting the operationalism which forms the framework of Wittgenstein's later writings, we do not, however, suppose that we have detracted in any way from the importance of the particular analyses of particular philosophical problems which form their primary content.

I

Among the philosophical problems Wittgenstein attempted to dissolve is the "problem of other

minds." One aspect of this hoary problem is the question: What justification, if any, can be given for the claim that one can tell, on the basis of someone's behavior, that he is in a certain mental state? To this question, the sceptic answers: No good justification at all. Among the major motivations of the later Wittgenstein's treatment of philosophical psychology is that of showing that this answer rests on a misconception and is *logically* incoherent.

Characteristically, philosophic sceptics have argued in the following way. It is assumed as a premiss that there are no logical or conceptual relations between propositions about mental states and propositions about behavior in virtue of which propositions asserting that a person behaves in a certain way provide support, grounds, or justification for ascribing the mental states to that person. From this, the sceptic deduces that he has no compelling reason for supposing that any person other than himself is ever truly said to feel pains, draw inferences, have motives, etc. For, while his first-hand knowledge of the occurrence of such mental events is of necessity limited to his own case, it is entailed by the premiss just cited that application of mental predicates to others must depend upon logically fallible inferences. Furthermore, attempts to base such inferences on analogies and correlations fall short of convincing justifications.

Various replies have been made to this argument which do not directly depend upon contesting the truth of the premiss. For example, it is sometimes claimed that, at least in some cases, no *inference*

¹ This work was supported in part by the U.S. Army, Navy, and Air Force under Contract DA 36-039-AMC-03200(E); in part by the National Science Foundation (Grant GP-2495), the National Institutes of Health (Grant MH-04737-04), the National Aeronautics and Space Administration (Ns G-496), the U.S. Air Force (ESD Contract AF 19 (628)-2487), the National Institute of Mental Health (Grant MPM 17, 760); and, in addition, by a University of California Faculty Fellowship.

² In making references to Part I of Ludwig Wittgenstein's *Philosophical Investigations* (New York, 1953), cited here as PI, we shall give section numbers, e.g. (PI, § 13), to Part II, we shall give page numbers, e.g. (PI, p. 226). In referring to his *The Blue and Brown Books* (New York, 1958), cited here as BB, we give page numbers. References to his *Remarks on the Foundations of Mathematics* (New York, 1956), cited here as RFM, will include both part and section numbers, e.g. (RFM, II, § 26).

from behavior to mental states is at issue in psychological ascriptions. Thus, we sometimes *see* that someone is in pain, and in these cases, we cannot be properly said to *infer* that he is in pain. However, the sceptic might maintain against this argument that it begs the question. For the essential issue is whether anyone is *justified* in claiming to see that another is in pain. Now a physicist, looking at cloud-chamber tracks, may be justified in claiming to see that a charged particle has passed through the chamber. That is because in this case there is justification for the claim that certain sorts of tracks show the presence and motion of particles. The physicist can explain not only how he is able to detect particles, but also why the methods he uses are methods of detecting *particles*. Correspondingly, the sceptic can argue that what is required in the case of another's pain is some justification for the claim that, by observing a person's behavior, one can *see* that he is in *pain*.

Wittgenstein's way of dealing with the sceptic is to attack his premiss by trying to show that there do exist conceptual relations between statements about behavior and statements about mental events, processes, and states. Hence, Wittgenstein argues that in many cases our knowledge of the mental states of some person rests upon something other than an observed empirical correlation or an analogical argument, viz. a conceptual or linguistic connection.

To hold that the sceptical premiss is false is *ipso facto* to commit oneself to some version of *logical behaviorism* where by "logical behaviorism" we mean the doctrine that there are logical or conceptual relations of the sort denied by the sceptical premiss.³ Which form of logical behaviorism one holds depends on the nature of the logical connection one claims obtains. The strongest form maintains that statements about mental states are translatable into statements about behavior. Wittgenstein, we shall argue, adopts a weaker version.

II

It is well known that Wittgenstein thought that philosophical problems generally arise out of misrepresentations and misinterpretations of ordinary language (PI, § 109, § 122, § 194). "Philosophy," he tells us, "is a fight against the fascination which forms of expression exert upon us" (BB, p. 27). Thus, Wittgenstein repeatedly warns us against being misled by superficial similarities between certain forms of expression (BB, p. 16) and tells us that, to avoid philosophical confusions, we must distinguish the "surface grammar" of sentences from their "depth grammar" (PI, § 11, § 664). For example, though the grammar of the sentence "A has a gold tooth" seems to differ in no essential respect from that of "A has a sore tooth," the apparent similarity masks important conceptual differences (BB, pp. 49, 53; PI, § 288-293). Overlooking these differences leads philosophers to suppose that there is a problem about our knowledge of other minds. It is the task of the Wittgensteinian philosopher to dissolve the problem by obtaining a clear view of the workings of pain language in this and other cases.

The Wittgensteinian method of philosophical therapy involves taking a certain view of language and of meaning. Throughout the *Investigations*, Wittgenstein emphasizes that "the speaking of language is part of an activity" (PI, § 23) and that if we are to see the radically different roles superficially similar expressions play, we must keep in mind the countless kinds of language-using activities or "language-games" in which we participate (BB, pp. 67-68).

It is clear that Wittgenstein thought that analyzing the meaning of a word involves exhibiting the role or use of the word in the various language-games in which it occurs. He even suggests that we "think of words as instruments characterized by their use . . ." (BB, p. 67).

³ Philosophers of Wittgensteinian persuasion have sometimes heatedly denied that the term "behaviorism" is correctly applied to the view that logical connections of the above sort exist. We do not feel that very much hangs on using the term "behaviorism" as we do, but we are prepared to give some justification for our terminology. "Behaviorism" is, in the first instance, a term applied to a school of psychologists whose interest was in placing constraints upon the conceptual equipment that might be employed in putative psychological explanations, but who were *not* particularly interested in the analysis of the mental vocabulary of ordinary language. The application of this label to a philosopher bent upon this latter task must therefore be, to some extent, analogical. Granted that there has been some tendency for the term "behaviorism" to be preempted, even in psychology, for the position held by such *radical* behaviorists as Watson and Skinner, who require that all psychological generalizations be defined over observables, insofar as C. L. Hull can be classified as a behaviorist, there does seem to be grounds for our classification. Hull's view, as we understand it, is that mental predicates are in no sense "eliminable" in favor of behavioral predicates, but that it is a condition upon their coherent employment that they be severally related to behavioral predicates and that some of these relations be logical rather than empirical—a view that is strikingly similar to the one we attribute to Wittgenstein. Cf. C. F. Hull, *Principles of Behavior* (New York, 1943).

This notion of analysis leads rather naturally to an operationalistic view of the meaning of certain sorts of predicates. For, in those cases where it makes sense to say of a predicate that one has determined that it applies, one of the central language-games that the fluent speaker has learned to play is that of making and reporting such determinations. Consider, for example, one of the language-games that imparts meaning to such words as "length," i.e., that of reporting the dimensions of physical objects. To describe this game, one would have to include an account of the procedures involved in measuring lengths; indeed, mastering (at least some of) those procedures would be an essential part of learning this game. "The meaning of the word 'length' is learnt among other things, by learning what it is to determine length" (PI, p. 225). As Wittgenstein comments about an analogous case, "Here the teaching of language is not explanation, but training" (PI, § 5). For Wittgenstein, "To understand a sentence means to understand a language." "To understand a language means to be master of a technique" (PI, § 199).

In short, part of being competent in the language-game played with "length" consists in the ability to arrive at the truth of such statements as "*x* is three feet long" by performing relevant operations with, e.g., rulers, range-finders, etc. A philosophic analysis of "length," insofar as it seeks to articulate the language-game played with that word, must thus refer to the operations which determine the applicability of length predicates. Finally, insofar as the meaning of the word is itself determined by the rules governing the language-games in which it occurs, a reference to these operations will be essential in characterizing the meaning of such predicates as "three feet long." It is in this manner that we are led to the view that the relevant operations for determining the applicability of a predicate are conceptually connected with the predicate.⁴

By parity of reasoning, we can see that to analyze such words as "pain," "motive," "dream," etc., will *inter alia* involve articulating the operations or observations in terms of which we determine that someone is in pain, or that he has such and such

a motive, or that he has dreamed, etc. (PI, p. 224). But clearly, such determinations are ultimately made on the basis of the behavior of the individual to whom the predicates are applied (taking behavior in the broad sense in which it includes verbal reports). Hence, for Wittgenstein, reference to the characteristic features of pain behavior on the basis of which we determine that someone is in pain is essential to the philosophical analysis of the word "pain" just as reference to the operations by which we determine the applicability of such predicates as "three feet long" is essential to the philosophical analysis of the word "length." In both cases, the relations are conceptual and the rule of language which articulates them is in that sense a rule of logic.

III

But what, specifically, is this logical connection which, according to Wittgenstein, is supposed to obtain between pain behavior and pain? Obviously, the connection is not that of simple entailment. It is evident that Wittgenstein did not think that some proposition to the effect that a person is screaming, wincing, groaning, or moaning could entail the proposition that the person is in pain. We know that Wittgenstein used the term "criterion" to mark this special connection, but we are in need of an explanation of this term.

We have already remarked that one of the central ideas in Wittgenstein's philosophy is that of a "language-game." Apparently Wittgenstein was passing a field on which a football game was being played when the idea occurred to him that "in language we play *games* with *words*."⁵ Since this analogy dominated so much of the later Wittgenstein's philosophical thinking, perhaps it would be well to begin the intricate task of explicating Wittgenstein's notion of criterion by considering some specific game.

Take basketball as an example. Since the object of the game is to score more points than one's opponents, there must be some way of telling if and when a team scores. Now there are various ways of telling that, say, a field goal has been scored. One might simply keep one's eyes on the

⁴ Cf. "Let us consider what we call an 'exact' explanation in contrast with this one. Perhaps something like drawing a chalk line round an area? Here it strikes us at once that the line has breadth. So a color-edge would be more exact. But has this exactness still got a function here: isn't the engine idling? And remember too that we have not yet defined what is to count as overstepping this exact boundary; *how, with what instruments, it is to be established*" (PI, § 88, italics ours). Cf., also RFM, I, § 5.

⁵ Norman Malcolm, *Ludwig Wittgenstein: a Memoir* (Oxford, 1958), p. 65.

scoreboard and wait for two points to be registered. Sometimes one realizes that a field goal has been scored on the basis of the reactions of the crowd. But these are, at best, indirect ways of telling, for if we use them we are relying on someone else: the score-keeper or other spectators. Obviously, not every way of telling is, in that sense, indirect; and anyone who is at all familiar with the game knows that, generally, one *sees* that a field goal has been scored in seeing the ball shot or tipped through the hoop. And if a philosopher asks, "Why does the fact that the ball went through the basket show that a field goal has been scored?" a natural reply would be, "That is what the rules of the game say; that is the way the game is played." The ball going through the basket satisfies a *criterion* for scoring a field goal.

Notice that though the relation between a criterion and that of which it is a criterion is a logical or conceptual one, the fact that the ball goes through the hoop does not entail that a field goal has been scored. First, the ball must be "in play" for it to be possible to score a field goal by tossing the ball through the basket. Second, even if the ball drops through the hoop when "in play," it need not follow that a field goal has been scored, for the rules of basketball do not cover all imaginable situations. Suppose, for example, that a player takes a long two-handed shot and that the ball suddenly reverses its direction, and after soaring and dipping through the air like a swallow in flight, gracefully drops through the player's own basket only to change into a bat, which immediately entangles itself in the net. What do the rules say about that?

An analogous situation would arise, in the case of a "language-game," if what seemed to be a chair suddenly disappeared, reappeared, and, in general, behaved in a fantastic manner. Wittgenstein's comment on this type of situation is:

Have you rules ready for such cases—rules saying whether one may use the word "chair" to include this kind of thing? But do we miss them when we use the word "chair"; and are we to say that we do not really attach any meaning to this word, because we are not equipped with rules for every possible application of it? (PI, § 80)

For Wittgenstein, a sign "is in order—if, under normal circumstances it fulfils its purpose." (PI, § 87).

It is only in normal cases that the use of a word is clearly prescribed; we know, are in no doubt, what to say in this or that case. The more abnormal the case, the more doubtful it becomes what we are to say. (PI, § 142)

Let us now try to make out Wittgenstein's distinction between *criterion* and *symptom*, again utilizing the example of basketball. Suppose that, while a game is in progress, a spectator leaves his seat. Though he is unable to see the playing court, he might realize that the home team had scored a field goal on the basis of a symptom—say, the distinctive roar of the crowd—which he had observed to be correlated with home-team field goals. This correlation, according to Wittgenstein, would have to be established *via* criteria, say, by noting the sound of the cheering when the home team shot the ball through the basket. Thus, a symptom is "a phenomenon of which experience has taught us that it coincided, in some way or other, with the phenomenon which is our defining criterion" (BB, p. 25). Though both symptoms and criteria are cited in answer to the question, "How do you know that so-and-so is the case?" (BB, p. 24), symptoms, unlike criteria, are discovered through experience or observation: that something is a symptom is not given by the rules of the "language-game" (not deducible from the rules alone). However, to say of a statement that it expresses a symptom is to say something about the relation between the statement and the rules, viz., that it is not derivable from them. Hence, Wittgenstein once claimed that "whereas 'When it rains the pavement gets wet' is not a grammatical statement at all, if we say 'The fact that the pavement is wet is a *symptom* that it has been raining' this statement is 'a matter of grammar'."⁶ Furthermore, giving the criterion for (e.g.) another's having a toothache "is to give a grammatical explanation about the word 'toothache' and, in this sense, an explanation concerning the meaning of the word 'toothache'" (BB, p. 24). However, given that there is this important difference between criteria and symptoms, the fact remains that Wittgenstein considered both symptoms and criteria as "evidences" (BB, p. 51).

Other salient features of criteria can be illuminated by exploiting our illustrative example. Consider Wittgenstein's claim that "in different circumstances we apply different criteria for a person's reading" (PI, § 164). It is clear that in

⁶ G. E. Moore, "Wittgenstein's Lectures in 1930-33," *Philosophical Papers* (London, 1959), pp. 266-267.

different circumstances we apply different criteria for a person's scoring a field goal. For example, the question whether a player scored a field goal may arise even though the ball went nowhere near the basket: in a "goal-tending" situation, the question will have to be decided on the basis of whether the ball had started its descent before the defensive player had deflected it. According to the rules it would be a decisive reason for not awarding a field goal that the ball had not reached its apogee when it was blocked.

One can now see that to claim that X is a criterion of T is not to claim that the presence, occurrence, existence, etc., of X is a necessary condition of the applicability of T , and it is not to claim that the presence, occurrence, existence, etc., of X is a sufficient condition of T , although if X is a criterion of T , it may be the case that X is a necessary or a sufficient condition of T .

Again, consider the tendency of Wittgenstein, noted by Albritton,⁷ to write as if X (a criterion of T) just is T or is what is called ' T ' in certain circumstances. We can understand a philosopher's wanting to say that shooting the ball through the basket in the appropriate situation just is scoring a field goal or is what we call "scoring a field goal."

Consider now the following passage from the *Investigations* (§ 376) which suggests a kind of test for "non-criterionhood":

When I say the ABC to myself, what is the criterion of my doing the same as someone else who silently repeats it to himself? It might be found that the same thing took place in my larynx and in his. (And similarly when we both think of the same thing, wish the same, and so on.) But then did we learn the use of the words: "to say such-and-such to oneself" by someone's pointing to a process in the larynx or the brain?

Obviously not. Hence, Wittgenstein suggests, something taking place in the larynx cannot be the criterion. The rationale behind this "test" seems to be this: For the teaching of a particular predicate ' T ' to be successful, the pupil must learn the rules for the use of ' T ' and hence must learn the criteria for ' T ' if there are such criteria. Thus,

⁷ Rogers Albritton, "On Wittgenstein's Use of the Term 'Criterion'," *Journal of Philosophy*, vol. 56 (1959), pp. 851-854.

⁸ Note Wittgenstein's suggestion that we can "give the phrase 'unconscious pain' sense by fixing experiential criteria for the case in which a man has pain and doesn't know it" (BB, p. 55). Cf., also: "If however we do use the expression 'the thought takes place in the head,' we have given this expression its meaning by describing the experience which would justify the hypothesis that the thought takes place in our heads, by describing the experience which we wish to call observing thought in our brain" (BB, p. 8).

⁹ Adopted by the eleventh General International Conference on Weights and Measures in the fall of 1960.

if the teaching could be entirely successful without one learning that X is something on the basis of which one tells that ' T ' applies, X cannot be a criterion of T . For example, since a person could be taught what "field goal" means without learning that one can generally tell that the home team has scored a field goal by noting the roar of the home crowd, the roar of the home crowd cannot be a criterion of field goals.

Finally, let us examine the principle, which Wittgenstein appears to maintain, that any change of criteria of X involves changing the concept of X . In the *Investigations*, Wittgenstein makes the puzzling claim:

There is one thing of which one can say neither that it is one metre long, nor that it is not one metre long, and that is the standard metre in Paris.—But this is, of course, not to ascribe any extraordinary property to it, but only to mark its peculiar role in the language-game of measuring with a metre-rule.—Let us imagine samples of colour being preserved in Paris like the standard metre. We define: "Sepia" means the colour of the standard sepia which is there kept hermetically sealed. Then it will make no sense to say of this sample either that it is of this colour or that it is not. (PI, § 50)

Wittgenstein evidently is maintaining not only that the senses of the predicates " x is one meter long" and " x is sepia" are given by the operations which determine the applicability of the respective predicates (the operations of comparing objects in certain ways with the respective standards),⁸ but also that these operations cannot be performed on the standards themselves and hence neither standard can be said to be an instance of either the predicate for which it is a standard or of its negation. (Cf., "A thing cannot be at the same time the measure and the thing measured" [RFM, I, § 40, notes].)

Wittgenstein would undoubtedly allow that we might introduce a new language-game in which "meter" is defined in terms of the wave length of the spectral line of the element krypton of atomic weight 86.⁹ In this language-game, where such highly accurate and complex measuring devices as the interferometer are required, the standard meter does not have any privileged position: it, too, can

be measured and "represented." In this language-game, the standard meter is or is not a meter. But here, Wittgenstein would evidently distinguish two senses of the term "meter." Obviously what is a meter in one language-game need not be a meter in the other. Thus, Wittgenstein's view seems to be that by introducing a new criterion for something's being a meter long, we have introduced a new language-game, a new sense of the term "meter," and a new concept of meter. Such a position is indicated by Wittgenstein's comment:

We can speak of measurements of time in which there is a different, and as we should say a greater, exactness than in the measurement of time by a pocket-watch; in which the words "to set the clock to the exact time" have a different, though related meaning. . . . (PI, § 88)

Returning to our basketball analogy, suppose that the National Collegiate Athletic Association ruled that, henceforth, a player can score a field goal by pushing the ball *upward* through the basket. Obviously, this would involve changing the rules of basketball. And to some extent, by introducing this new criterion, the rules governing the use or "grammar" of the term "field goal" would be altered. To put it somewhat dramatically (in the Wittgensteinian style), a new *essence* of field goal would be created. (Cf. "The mathematician creates *essence*" [RFM, I, § 32].) For Wittgenstein, not only is it the case that the criteria we use "give our words their common meanings" (BB, p. 57) and that to explain the criteria we use is to explain the meanings of words (BB, p. 24), but also it is the case that to introduce a new criterion of X is to define a new concept of X .¹⁰

In summary, we can roughly and schematically characterize Wittgenstein's notion of criterion in the following way: X is a criterion of Y in situations of type S if the very meaning or definition of ' Y ' (or, as Wittgenstein might have put it, if the "grammatical" rules for the use of ' Y ')¹¹ justify the claim that one can recognize, see, detect, or determine the applicability of ' Y ' on the basis of X in *normal* situations of type S . Hence, if the above relation obtains between X and Y , and if someone admits that X but denies Y , the burden of proof is upon him to show that something is abnormal in the situation. In a normal situation,

the problem of gathering evidence which justifies concluding Y from X simply does not arise.

IV

The following passage occurs in the *Blue Book* (p. 24):

When we learnt the use of the phrase "so-and-so has toothache" we were pointed out certain kinds of behavior of those who were said to have toothache. As an instance of these kinds of behavior let us take holding your cheek. Suppose that by observation I found that in certain cases whenever these first criteria told me a person had toothache, a red patch appeared on the person's cheek. Supposing I now said to someone "I see A has toothache, he's got a red patch on his cheek." He may ask me "How do you know A has toothache when you see a red patch?" I would then point out that certain phenomena had always coincided with the appearance of the red patch.

Now one may go on and ask: "How do you know that he has got toothache when he holds his cheek?" The answer to this might be, "I say, *he* has toothache when he holds his cheek because I hold my cheek when I have toothache." But what if we went on asking:—"And why do you suppose that toothache corresponds to his holding his cheek just because your toothache corresponds to your holding your cheek?" You will be at a loss to answer this question, and find that here we strike rock bottom, that is we have come down to conventions.

It would seem that, on Wittgenstein's view, empirical justification of the claim to see, recognize, or know that such and such is the case *on the basis of some observable feature or state of affairs*, would have to rest upon inductions from observed correlations, so that, if a person claims that Y is the case on the grounds that X is the case, in answer to the question "Why does the fact that X show that Y ?" he would have to cite either conventions or observed correlations linking X and Y . Thus, Wittgenstein appears to be arguing that the possibility of ever inferring a person's toothache from his behavior requires the existence of a criterion of toothache that can sometimes be observed to obtain. A generalized form of this argument leads to the conclusion that "an 'inner process' stands in need of outward criteria" (PI, § 580).

As an illustration of Wittgenstein's reasoning,

¹⁰ RFM, II, § 24; III, § 29; and I, Appendix I, § 15-16. See also C. S. Chihara "Mathematical Discovery and Concept Formation," *The Philosophical Review*, vol. 72 (1963), pp. 17-34.

¹¹ Cf., "The person of whom we say 'he has pain' is, *by the rules of the game*, the person who cries, contorts his face, etc." (BB, p. 68, italics ours).

consider the following example: It appears to be the case that the measurement of the alcohol content of the blood affords a reasonably reliable index of intoxication. On the basis of this empirical information, we may sometimes justify the claim that *X* is intoxicated by showing that the alcohol content of his blood is higher than some specified percentage. But now consider the justification of the claim that blood-alcohol is in fact an index of intoxication. On Wittgenstein's view, the justification of *this* claim must rest ultimately upon correlating cases of intoxication with determinations of high blood-alcohol content. But, the observations required for this correlation could be made only if there exist independent techniques for identifying each of the correlated items. In any particular case, these independent techniques may themselves be based upon further empirical correlations; we might justify the claim that the blood-alcohol content is high by appealing to some previously established correlation between the presence of blood-alcohol and some test result. But ultimately according to Wittgenstein, we must come upon identifying techniques based not upon further empirical correlations, but rather upon definitions or conventions which determine criteria for applying the relevant predicates. This is why Wittgenstein can say that a symptom is "a phenomenon of which experience has taught us that it coincided, in some way or other with the phenomenon which is our defining criterion" (BB, p. 25).

A similar argument has recently been given by Sidney Shoemaker who writes:

If we know psychological facts about other persons at all, we know them on the basis of their behavior (including, of course, their verbal behavior). Sometimes we make psychological statements about other persons on the basis of bodily or behavioral facts that are only contingently related to the psychological facts for which we accept them as evidence. But we do this only because we have discovered, or think we have discovered, empirical correlations between physical (bodily and behavioral) facts of a certain kind and psychological facts of a certain kind. And if *all* relations between physical and psychological facts were contingent, it would be impossible for us to discover such correlations. . . . Unless some relationships between physical and psychological states are not contingent, and can be known prior to the discovery of empirical correlations, we cannot have even indirect inductive evidence for the truth of psychological statements about other persons, and

cannot know such statements to be true or even probably true.¹²

Malcolm argues in a similar manner in *Dreaming*.¹³

Of course, Wittgenstein did not claim that all predicates presuppose criteria of applicability. For example, Wittgenstein probably did not think that we, in general, see, tell, determine, or know that something is red on the basis of either a criterion or a symptom. The relevant difference between ascriptions of "red" and third-person ascriptions of "pain" is that we generally see, recognize, determine, or know that another is in pain on the basis of something which is not the pain itself (as for example, behavior and circumstances) whereas, if it made any sense at all to say we generally see, recognize, etc., that an object is red on the basis of something, what could this something be other than just the object's redness? But Wittgenstein's use of the term "criterion" seems to preclude redness being a criterion of redness. If someone asks "How do you know or tell that an object is red?" it would not, in general, do to answer "By its redness." (Cf. Wittgenstein's comment "How do I know that this color is red?—It would be an answer to say: 'I have learnt English'" [PI, § 381].) Evidently, some color predicates and, more generally, what are sometimes called "sense datum" predicates (those that can be known to apply—as some philosophers put it—*immediately*), do not fall within the domain of arguments of the above type. But the predicates with which we assign "inner states" to another person are not of this sort. One recognizes that another is in a certain mental state, *Y*, on the basis of something, say, *X*. Now it is assumed that *X* must be either a criterion or symptom of *Y*. If *X* is a symptom, *X* must be known to be correlated with *Y*, and we may then inquire into the way in which this correlation was established. Again, *X* must have been observed to be correlated with a criterion of *Y* or with a symptom, *X*₁, of *Y*. On the second alternative, we may inquire into the basis for holding that *X*₁ is a symptom of *Y*. . . . Such a chain may go on for any distance you like, but it cannot go on indefinitely. That is, at some point, we must come to a criterion of *Y*. But once this conclusion has been accepted, there appears to be no reasonable non-sceptical alternative to Wittgenstein's logical behaviorism, for if "inner" states require "outward" criteria, behavioral criteria are the only plausible candidates.

¹² Sidney Shoemaker, *Self-knowledge and Self-identity* (Ithaca, 1963), pp. 167–168.

¹³ Norman Malcolm, *Dreaming* (London, 1959), pp. 60–61.

V

As a refutation of scepticism, the above argument certainly will not do; for, at best, it supports Wittgenstein's position only on the assumption that the sceptic is not right. That is, it demonstrates that there must be criteria for psychological predicates by assuming that such predicates are sometimes applied justifiably. A sceptic who accepts the argument of Section IV could maintain his position only by allowing that no one could have any idea of what would show or even indicate that another is in pain, having a dream, thinking, etc. In this section we shall show how Wittgenstein argues that that move would lead the sceptic to the absurd conclusion that it must be impossible to teach the meaning of these psychological predicates.

"What would it be like if human beings showed no outward signs of pain (did not groan, grimace, etc.)? Then it would be impossible to teach a child the use of the word 'toothache'" (PI, § 257). For just imagine trying to teach a child the meaning of the term "toothache," say, on the supposition that there is absolutely no way of telling whether the child—or anyone else for that matter—is actually in pain. How would one go about it, if one had no reason for believing that gross damage to the body causes pain or that crying out, wincing, and the like indicate pain? ("How could I even have come by the idea of another's experience if there is no possibility of any evidence for it?" [BB, p. 46; cf. also BB, p. 48].)

Again, what would show us that the child had grasped the teaching? If anything would, the argument of Section IV requires that there be a criterion of having succeeded in teaching the child. (As Wittgenstein says of an analogous case, "If I speak of communicating a feeling to someone else, mustn't I in order to understand what I say know what I shall call the criterion of having succeeded in communicating?" (BB, p. 185).) But the only plausible criterion of this would be that the child applies the psychological predicates correctly (cf. PI, § 146); and since the sceptical position implies that there is no way of knowing if the child correctly applies such predicates, it would seem to follow that nothing could show or indicate that the child had learned what these terms mean.

We now have a basis for explicating the sense

of "logical" which is involved in the claim that scepticism is a logically incoherent doctrine. What Wittgenstein holds is not that "*P* and not-*P*" are strictly deducible from the sceptic's position, but rather that the sceptic's view presupposes a deviation from the rules for the use of key terms. In particular, Wittgenstein holds that if the sceptic were right, the preconditions for teaching the meaning of the mental predicates of our ordinary language could not be satisfied.¹⁴

We now see too the point to the insistence that the sceptic's position must incorporate an extraordinary and misleading use of mental predicates. The sceptic's view is logically incompatible with the operation of the ordinary language rules for the application of these terms, and these rules determine their meanings. (Cf. "What *we* do is to bring words back from their metaphysical to their everyday usage" [PI, § 116].) As Wittgenstein diagnoses the sceptic's view, the sceptic does not have in mind any criteria of third person ascriptions when he denies that he can know if anyone else has pains (cf. PI, § 272). The sceptic tempts us to picture the situation as involving "a barrier which doesn't allow one person to come closer to another's experience than to the point of observing his behavior"; but, according to Wittgenstein, "on looking closer we find that we can't apply the picture" (BB, p. 56); no clear meaning can be attached to the sceptic's claim: no sense can even be given the hypothesis that other people feel "pains," as the sceptic uses the term "pain." ("For how can I even make the hypothesis if it transcends all possible experience?" [BB, p. 48].) And if the sceptic says, "But if I suppose that someone has a pain, then I am simply supposing that he has just the same as I have so often had," Wittgenstein can reply:

That gets us no further. It is as if I were to say: "You surely know what 'It is 5 o'clock here' means; so you also know what 'It's 5 o'clock on the sun' means. It means simply that it is just the same time there as it is here when it is 5 o'clock."—The explanation by means of *identity* does not work here. For I know well enough that one can call 5 o'clock here and 5 o'clock there "the same time," but what I do not know is in what cases one is to speak of its being the same time here and there. (PI, § 350)

Thus, we can see how Wittgenstein supports his logical behaviorism: the argument in Section IV

¹⁴ Cf., "Before I judge that two images which I have are the same, I must recognize them as the same." . . . Only if I can express my recognition in some other way, and if it is possible for someone else to teach me that 'same' is the correct word here" (PI, § 378).

purports to show that the only plausible alternative to Wittgenstein's philosophical psychology is radical scepticism; and the argument in the present section rules out this alternative. For Wittgenstein, then, "the person of whom we say 'he has pains' is, by the rules of the game, the person who cries, contorts his face, etc.," (BB, p. 68).

Undoubtedly, there is much that philosophers find comforting and attractive in Wittgenstein's philosophical psychology, but there are also difficulties in the doctrine which mar its attractiveness. To some of these difficulties, we shall now turn.

VI

In this section, we shall consider some consequences of applying the views just discussed to the analysis of dreaming, and we shall attempt to show that the conclusions to which these views lead are counter-intuitive.

According to Wittgenstein, we are to understand the concept of dreaming in terms of the language-game(s) in which "dream" plays a role and, in particular, in terms of the language-game of dream telling. For, to master the use of the word "dream" is precisely to learn what it is to find out that someone has dreamed, to tell what someone has dreamed, to report one's own dreams, and so on. Passages in the *Investigations* (e.g., PI, pp. 184, 222-223) indicate that, for Wittgenstein, a criterion of someone's having dreamed is the dream report. On this analysis, sceptical doubts about dreams arise when we fail to appreciate the logical bond between statements about dreams and statements about dream reports. The sceptic treats the dream report as, at best, an empirical correlate of the occurrence of a dream: a symptom that is, at any event, no more reliable than the memory of the subject who reports the dream. But, according to Wittgenstein, once we have understood the criterial relation between dream reporting and dreaming, we see that "the question whether the dreamer's memory deceives him when he reports the dream after waking cannot arise . . ." (PI, p. 222). (Compare: "Once we understand the rules for playing chess, the question whether a player has won when he has achieved check-mate cannot arise.")

The rules articulating the criteria for applying the word "dream" determine a logical relation between dreaming and reporting dreams. Moreover, the set of such rules fixes the language-game

in which "dream" has its role and hence determines the meaning of the word.

It is important to notice that there are a number of *prima facie* objections to this analysis which, though perhaps not conclusive, supply grounds for questioning the doctrines which lead to it. Though we could perhaps learn to live with these objections were no other analyses available, when seen from the vantage point of an alternative theory they indicate deep troubles with Wittgenstein's views.

(1) Given that there exist no criteria for first person applications of many psychological predicates ("pain," "wish," or the like) it is unclear how the first person aspects of the game played with these predicates are to be described. Wittgenstein does not appear to present a coherent account of the behavior of predicates whose applicability is not determined by criteria. On the other hand, the attempt to characterize "I dreamt" as criterion-governed leads immediately to absurdities. Thus, in Malcolm's *Dreaming* it is suggested that:

If a man wakes up with the impression of having seen and done various things, and if it is known that he did not see and do those things, then it is known that he dreamt them. . . . When he says "I dreamt so and so" he implies, first, that it seemed to him on waking up as if the so and so had occurred and second, that the so and so did not occur. (p. 66)

That this is an incredibly counter-intuitive analysis of our concept of dreaming hardly needs mentioning. We ask the reader to consider the following example: A person, from time to time, gets the strange feeling that, shortly before, he had seen and heard his father commanding him to come home. One morning he wakes with this feeling, knowing full well that his father is dead. Now we are asked by Malcolm to believe that the person *must have dreamt* that he saw and heard his father: supposedly, it would be logically absurd for the person to claim to have this feeling and deny that he had dreamt it!

(2) Wittgenstein's view appears to entail that no sense can be made of such statements as "Jones totally forgot the dream he had last night," since we seem to have no criteria for determining the truth of such a statement. (We have in mind the case in which Jones is totally unable to remember having dreamed and no behavioral manifestations of dreaming were exhibited.) It is sometimes denied that observations of what people ordinarily say are relevant to a description of ordinary language.

But, insofar as statements about what we would say are susceptible to empirical disconfirmation, the claim that we would feel hesitation about saying that someone completely forgot his dream appears to be just false.¹⁵

(3) The Wittgensteinian method of counting concepts is certainly not an intuitive one. Consider Malcolm's analysis of dreaming again. Malcolm realizes that sometimes, on the basis of a person's behavior during sleep, we say that he had a dream, even though he is unable to recall a dream upon awaking. But, in such cases, Malcolm claims, "our words . . . have no clear sense" (*Dreaming*, p. 62). On the other hand, Malcolm admits that there is a *sense* of the term "nightmare" where behavior during sleep is the criterion. However, a different concept of dreaming is supposedly involved in this case. An analogous situation is treated in the *Blue Book* (p. 63), where Wittgenstein writes:

If a man tries to obey the order "Point to your eye," he may do many different things, and there are many different criteria which he will accept for having pointed to his eye. If these criteria, as they usually do, coincide, I may use them alternately and in different combinations to show me that I have touched my eye. If they don't coincide, I shall have to distinguish between different senses of the phrase "I touch my eye" or "I move my finger towards my eye."

Following this suggestion of Wittgenstein, Malcolm distinguishes not only different senses of the term "dream," but also different concepts of sleep—one based upon report, one based upon nonverbal behavior. But surely, this is an unnatural way of counting concepts. Compare Malcolm's two concepts of sleep with a case where it really does seem natural to say that a special concept of sleep has been employed, viz., where we say of a hibernating bear that it sleeps through the winter.

(4) As Malcolm points out, the language-game *now* played with "dream" seems to exhibit no criteria which would enable one to determine the precise duration of dreams. Hence, it would seem to follow (as Malcolm has noticed) that scientists who have attempted to answer such questions as, "How long do dreams last?" are involved in conceptual confusions rather than empirical determinations. For such questions cannot be answered without adopting criteria for ascribing the relevant properties to dreams. But since, on Wittgenstein's view, to adopt such new criteria for the use of a word is, to that extent, to change its meaning, it follows that the concept of "dream" that such researchers employ is not the ordinary concept and hence that the measurements they effect are not, strictly speaking, measurements of *dreams*.¹⁶ The notion that adopting any test for dreaming which arrives at features of dreams not determinable from the dream report thereby alters the concept of a dream seems to run counter to our intuitions about the goals of psychological research. It is not immediately obvious that the psychologist who says he has found a method of measuring the duration of dreams *ipso facto* commits the fallacy of ambiguity.¹⁷

(5) Consider the fact that such measures as EEG, eye-movements and "dream-behavior" (murmuring, tossing, etc., during sleep) correlate reasonably reliably with one another and dream reports. The relation between, say, EEG and dream reports is clearly not criterial; no one holds that EEG is a criterion of dream reports. It would seem then that, on Wittgenstein's view, EEG provides us with, at best, a symptom of positive dream reports; and symptoms are supposedly discovered by observing co-occurrences. The difficulty, however, is that this makes it unclear how the expectation that such a correlation must obtain could have been a rational expectation even *before* the correla-

¹⁵ Thus consider the following: "Up until the night I opened the door, I remembered my dreams. Soon after, I ceased to recall them. I still dreamed, but my waking consciousness concealed from itself what sleep revealed. If the recurrent nightmare of the iron fence awoke me, I recognized it. But if any other nightmare broke my sleep, I forgot what it was about by morning. And of all the other dreams I had during the night I remembered nothing" (Windham, D., "Myopia," *The New Yorker*, July 13, 1963).

¹⁶ In *Dreaming*, Malcolm gives a number of arguments, not to be found in Wittgenstein's published writings, for the position that psychologists attempting to discover methods of measuring the duration of dreams must be using the term "dream" in a misleading and extraordinary way. For a reply to these arguments, see C. S. Chihara "What Dreams are Made On" forthcoming in *Theoria*. See also H. Putnam's criticism of Malcolm, "Dreaming and 'Depth Grammar,'" *Analytical Philosophy*, ed. by R. J. Butler (Oxford, 1962), pp. 211-235.

¹⁷ The implausibility of this view is even more striking when Wittgenstein applies it in his philosophy of mathematics to arrive at the conclusion that every new theorem about a concept alters the concept or introduces a new concept. When the notion of conceptual change is allowed to degenerate this far, it is not easy to see that anything rides on the claim that a conceptual change has taken place. Cf. C. S. Chihara, "Mathematical Discovery and Concept Formation," *The Philosophical Review*, vol. 72 (1963), pp. 17-34.

tion was experimentally confirmed. One cannot have an inductive generalization over no observations; nor, in this case, was any higher level "covering law" used to infer the probability of a correlation between EEG and dream reports. Given Wittgenstein's analysis of the concept of dreaming, not only do the researches of psychologists into the nature of dreams appear mysterious, but even the expectations, based upon these researches, seem somewhat irrational.

The difficulties we have mentioned are not peculiar to the Wittgensteinian analysis of dreams. Most of them have counterparts in the analyses of sensation, perception, intention, etc. Whether or not these difficulties can be obviated, in some way, noticing them provides a motive for re-examining the deeper doctrines upon which Wittgensteinian analyses of psychological terms are based.

VIII

The Wittgensteinian argument of Section IV rests on the premiss that if we are justified in claiming that one can tell, recognize, see, or determine that '*Y*' applies on the basis of the presence of *X*, then either *X* is a criterion of *Y* or observations have shown that *X* is correlated with *Y*. Wittgenstein does not present any justification for this premiss in his published writings. Evidently, some philosophers find it self-evident and hence in need of no justification. We, on the other hand, far from finding this premiss self-evident, believe it to be false. Consider: one standard instrument used in the detection of high-speed, charged particles is the Wilson cloud-chamber. According to present scientific theories, the formation of tiny, thin bands of fog on the glass surface of the instrument indicates the passage of charged particles through the chamber. It is obvious that the formation of these streaks is not a Wittgensteinian criterion of the presence and motion of these particles in the apparatus. That one can detect these charged particles and determine their paths by means of such devices is surely not, by any stretch of the imagination, a *conceptual* truth. C. T. R. Wilson did not learn what "path of a charged particle" means by having the cloud-chamber explained to him: he *discovered* the method, and the discovery was contingent upon recognizing the empirical fact that ions could act as centers of condensation in a supersaturated vapor. Hence, applying Wittgenstein's own test for non-criterionhood (see above), the formation

of a cloud-chamber track cannot be a criterion of the presence and motion of charged particles.

It is equally clear that the basis for taking these streaks as indicators of the paths of the particles is not observed *correlations* between streaks and some criterion of motion of charged particles. (What criterion for determining the path of an electron could Wilson have used to establish such correlations?) Rather, scientists were able to give compelling explanations of the formation of the streaks on the hypothesis that high-velocity, charged particles were passing through the chamber; on this hypothesis, further predictions were made, tested, and confirmed; no other equally plausible explanation is available; and so forth.

Such cases suggest that Wittgenstein failed to consider all the possible types of answers to the question, "What is the justification for the claim that one can tell, recognize, or determine that '*Y*' applies on the basis of the presence of *X*?" For, where '*Y*' is the predicate "is the path of a high-velocity particle," *X* need not have the form of either a criterion or a correlate.

Wittgensteinians may be tempted to argue that cloud-chamber tracks really are criteria, or symptoms observed to be correlated with criteria, of the paths of charged particles. To obviate this type of counter, we wish to stress that the example just given is by no means idiosyncratic. The reader who is not satisfied with it will easily construct others from the history of science. What is at issue is the possibility of a type of justification which consists in neither the appeal to criteria nor the appeal to observed correlations. If the Wittgensteinian argument we have been considering is to be compelling, some grounds must be given for the exhaustiveness of these types of justification. This, it would seem, Wittgenstein has failed to do.

It is worth noticing that a plausible solution to the problem raised in VI. 5 can be given if we consider experiments with dreams and EEG to be analogous to the cloud-chamber case. That is, we can see how it could be the case that the correlation of EEG with dream reports was anticipated prior to observation. The dream report was taken by the experiments to be an indicator of a psychological event occurring prior to it. Given considerations about the relation of cortical to psychological events, and given also the theory of EEG, it was predicted that the EEG should provide an index of the occurrence of dreams. From the hypothesis that dream reports and EEG readings

are both indices of the same psychological events, it could be deduced that they ought to be reliably correlated with one another, and this deduction in fact proved to be correct.

This situation is not at all unusual in the case of explanations based upon theoretical inferences to events underlying observable syndromes. As Meehl and Cronbach have pointed out, in such cases the validity of the "criterion" is often nearly as much at issue as the validity of the indices to be correlated with it.¹⁸ The successful prediction of the correlation on the basis of the postulation of a common etiology is taken both as evidence for the existence of the cause and as indicating the validity of each of the correlates as an index of its presence.

In this kind of case, the justification of existential statements is thus identical neither with an appeal to criteria nor with an appeal to symptoms. Such justifications depend rather on appeals to the simplicity, plausibility, and predictive adequacy of an explanatory system as a whole, so that it is incorrect to say that relations between statements which are mediated by such explanations are either logical in Wittgenstein's sense or contingent in the sense in which this term suggests simple correlation.

It cannot be stressed too often that there exist patterns of justificatory argument which are not happily identified either with appeals to symptoms or with appeals to criteria, and which do not in any obvious way rest upon such appeals. In these arguments, existential claims about states, events, and processes, which are *not* directly observable are susceptible of justification despite the fact that no *logical* relation obtains between the predicates ascribing such states and predicates whose applicability *can* be directly observed. There is a temptation to hold that in such cases there *must* be a criterion, that there must be some set of possible observations which would settle *for sure* whether the theoretical predicate applies. But we succumb to this temptation at the price of postulating stipulative definitions and conceptual alterations which fail to correspond to anything we can discover in the course of empirical arguments. The counter-intuitive features of philosophic analyses based on the assumption that there must be criteria are thus not the consequences of a profound

methodological insight, but rather a projection of an inadequate philosophical theory of justification.

IX

It might be replied that the above examples do not constitute counter-instances to Wittgenstein's criterion-correlation premiss since Wittgenstein may have intended his principle to be applicable only in the case of ordinary language terms which, so it might seem, do not function within the framework of a theory. It is perhaps possible to have indicators that are neither criteria nor symptoms of such highly theoretical entities as electrons and positrons, but the terms used by ordinary people in everyday life are obviously (?) in a different category. (Notice that Wittgenstein considers "making scientific hypotheses and theories" a different "game" from such "language-games" as "describing an event" and "describing an immediate experience" [BB, pp. 67-68; Cf. PI, § 23].) Hence, Wittgenstein might argue, it is only in the case of ordinary language terms that the demand for criteria is necessary.

Once one perceives the presuppositions of Wittgenstein's demand for criteria, however, it becomes evident that alternatives to Wittgenstein's analyses of ordinary language mental terms should at least be explored. Perhaps, what we all learn in learning what such terms as "pain" and "dream" mean are not criterial connections which map these terms severally onto characteristic patterns of behavior. We may instead form complex conceptual connections which interrelate a wide variety of mental states. It is to such a conceptual system that we appeal when we attempt to explain someone's behavior by reference to his motives, intentions, beliefs, desires, or sensations. In other words, in learning the language, we develop a number of intricately interrelated "mental concepts" which we use in dealing with, coming to terms with, understanding, explaining, interpreting, etc., the behavior of other human beings (as well as our own). In the course of acquiring these mental concepts we develop a variety of beliefs involving them. Such beliefs result in a wide range of expectations about how people are likely to behave. Since only a portion of these beliefs are confirmed in the normal course, these beliefs and the con-

¹⁸ P. M. Meehl and H. J. Cronbach, "Construct Validity in Psychological Tests," *Minnesota Studies in the Philosophy of Science*, vol. I, ed. by H. Feigl and M. Scriven (Minneapolis, 1956), pp. 174-204. We have followed Meehl and Cronbach's usage of the terms "reliability" and "validity" so that *reliability* is a measure of the correlation between criteria while *validity* is a measure of the correlation between a criterion and the construct whose presence it is supposed to indicate.

ceptual systems which they articulate are both subject to correction and alteration as the consequence of our constant interaction with other people.

On this view, our success in accounting for the behavior on the basis of which mental predicates are applied might properly be thought of as supplying *evidence* for the existence of the mental processes we postulate. It does so by attesting to the adequacy of the conceptual system in terms of which the processes are understood. The behavior would be, in that sense, analogous to the cloud-chamber track on the basis of which we detect the presence and motion of charged particles. Correspondingly, the conceptual system is analogous to the physical *theory* in which the properties of these particles are formulated.

If something like this should be correct, it would be possible, at least in theory, to reconstruct and describe the conceptual system involved and then to obtain some confirmation that the putative system is in fact employed by English speakers. For example, confirmation might come *via* the usual methods of "reading off" the conceptual relation in the putative system and *matching them* against the linguistic intuitions of native speakers. Thus, given that a particular conceptual system is being employed, certain statements should strike native speakers as nonsensical, others should seem necessarily true, others should seem ambiguous, others empirically false, and so on, all of which would be testable.

To maintain that there are no criterial connections between pains and behavior does not commit us to holding that the fact that people often feel *pains* when they cry out is *just* a contingent fact (in the sense in which it is just a contingent fact that most of the books in my library are unread). The belief that other people feel pains is not gratuitous even on the view that there are no criteria of pains. On the contrary, it provides the only plausible explanation of the facts I know about the way that they behave in and *vis à vis* the sorts of situations I find painful. These facts are, of course, enormously complex. The "pain syndrome" includes not only correlations between varieties of overt behaviors but also more subtle relations between pain and motivations, utilities, desires, and so on. Moreover, I confidently expect that there must exist reliable members of this syndrome other than the ones with which I am currently familiar. I am in need of an explanation of the reliability and fruitfulness of this syndrome, an

explanation which reference to the occurrence of pains supplies. Here, as elsewhere, an "outer" syndrome stands in need of an inner process.

Thus, it is at least conceivable that a non-Wittgensteinian account ought to be given of the way children learn the mental predicates. (It is, at any event, sufficient to notice that such an account *could* be given, that there exist alternatives to Wittgenstein's doctrine.) For example, if the concept of dreaming is *inter alia* that of an inner event which takes place during a definite stretch of "real" time, which causes such involuntary behavior as moaning and murmuring in one's sleep, tossing about, etc., and which is remembered when one correctly reports a dream, then there are a number of ways in which a child might be supposed to "get" this concept other than by learning criteria for the application of the word "dream." Perhaps it is true of many children that they learn what a dream is by being told that what they have just experienced was a dream. Perhaps it was also true of many children that, having grasped the notions of *imagining* and *sleep*, they learn what a dream is when they are told that dreaming is something like imagining in your sleep.

But does this imply that children learn what a dream is "from their own case?" If this is a logical rather than psychological question, the answer is "Not necessarily": a child who never dreamed, but who was very clever, might arrive at an understanding of what dreams are just on the basis of the sort of theoretical inference we have described above. For our notion of a dream is that of a mental event having various properties that are required in order to explain the characteristic features of the dream-behavior syndrome. For example, dreams occur during sleep, have duration, sometimes cause people who are sleeping to murmur or to toss, can be described in visual, auditory, or tactile terms, are sometimes remembered and sometimes not, are sometimes reported and sometimes not, sometimes prove frightening, sometimes are interrupted before they are finished, etc. But if these are the sorts of facts that characterize our concept of dream, then there seems to be nothing which would, in principle, prevent a child who never dreamed from arriving at this notion.

A similar story might be told about how such sensation terms as "pain" are learned and about the learning of such quasi-dispositionals as "having a motive." In each case, since the features that

we in fact attribute to these states, processes, or dispositions are just those features we know they must have if they are to fulfill their role in explanations of behavior, etiology, personality, etc., it would seem that there is nothing about them the child could not in principle learn by employing the pattern of inference we have described above, and hence nothing that he could in principle learn *only* by an analogy to his own case.

Now it might be argued that the alternative to Wittgenstein's position we have been sketching is highly implausible. For, if children do have to acquire the complicated conceptual system our theory requires to understand and use mental predicates, surely they would have to be taught this system. And the teaching would surely have to be terribly involved and complex. But as a matter of fact, children do not require any such teaching at all, and hence we should conclude that our alternative to Wittgenstein's criterion view is untenable.

The force of this argument, however, can to some extent be dispelled if we consider the child's acquisition of, e.g., the grammar of a natural language. It is clear that, by some process we are only now beginning to understand, a child, on the basis of a relatively short "exposure" to utterances in his language, develops capacities for producing and understanding "novel" sentences (sentences which he has never previously heard or seen). The exercise of these capacities, so far as we can tell, "involve" the use of an intricate system of linguistic rules of very considerable generality and complexity.¹⁹ That the child is not taught (in any ordinary sense) any such system of rules is undeniable. These capacities seem to develop naturally in the child in response to little more than contact with a relatively small number of sentences uttered in ordinary contexts in everyday life.²⁰ Granting for the moment that the apparent complexity of such systems of rules is not somehow an artifact of an unsatisfactory theory of language, the fact that the child develops these linguistic capacities shows that a corresponding "natural" development of a system of mental concepts may not, as a matter of brute fact, require the sort of explicit teaching a person needs to master, say, calculus or quantum physics.

X

It is easily seen that this unabashedly non-behavioristic view avoids each of the difficulties we raised regarding Wittgenstein's analyses of mental predicates. Thus, the asymmetry between first and third person uses of "dream" discussed in Section VI need not arise since there need be no criteria for "*X* dreamed," *whatever* value *X* takes: we do not have the special problem of characterizing the meaning of "I dreamed" since "dream" in this context means just what it means in third person contexts, viz., "a series of thoughts, images, or emotions occurring during sleep." Again, it is now clear why people find such remarks as "Jones totally forgot what and that he dreamed last night" perfectly sensible. It is even clear how such assertions might be confirmed. Suppose, for example, that there exists a neurological state α such that there is a very high correlation between the presence of α and such dream behavior as tossing in one's sleep, crying out in one's sleep, reporting dreams, and so on. Suppose, too that there exists some neurological state β such that whenever β occurs, experiences that the subject has had just prior to β are forgotten. Suppose, finally, that sometimes we observe sequences, α , β , and that such sequences are not followed by dream reports though the occurrences of α are accompanied by other characteristic dream behaviors. It seems clear that the reasonable thing to say in such a case is that the subject has dreamed and forgotten his dream. And since we have postulated no criterion for dreaming, but only a syndrome of dream behaviors each related to some inner psychological event, we need have no fear that, in saying what it is reasonable to say, we have changed the meaning of "dream." We leave it to the reader to verify that the other objections we raised against the Wittgensteinian analysis of "dream" also fail to apply to the present doctrine.

Thus, once we have abandoned the arguments for a criterial connection between statements about behavior and statements about psychological states, the question remains open whether applications of ordinary language psychological terms on the basis of observations of behavior ought not themselves be treated as theoretical inferences to underlying mental occurrences. The question

¹⁹ This point is susceptible of direct empirical ratification, for it can be demonstrated that in perceptual analysis, speech is analyzed into segments which correspond precisely to the segmentation assigned by a grammar.

²⁰ Cf. N. Chomsky's "A Review of Skinner's *Verbal Behavior*," reprinted in J. Fodor and J. Katz, *The Structure of Language* (Englewood Cliffs, 1964).

whether such statements as "He moaned because he was in pain" function to explain behavior by relating it to an assumed mental event cannot be settled simply by reference to ordinary linguistic usage. Answering this question requires broadly empirical investigations into the nature of thought and concept formation in normal human beings. What is at issue is the question of the role of theory construction and theoretical inference in thought and argument outside pure science. Psychological investigations indicate that much everyday conceptualization depends on the exploitation of theories and explanatory models in terms of which experience is integrated and understood.²¹ Such pre-scientific theories, far from being mere functionless "pictures," play an essential role in determining the sorts of perceptual and inductive expectations we form and the kind of arguments and explanations we accept. It thus seems *possible* that the correct view of the functioning of ordinary language mental predicates would assimilate apply-

ing them to the sorts of processes of theoretical inference operative in scientific psychological explanation. If this is correct, the primary difference between ordinary and scientific uses of psychological predicates would be just that the processes of inference which are made explicit in the latter case remain implicit in the former.

We can now see what should be said in reply to Wittgenstein's argument that the possibility of teaching a language rests upon the existence of criteria. Perhaps teaching a word would be impossible if it could not sometimes be determined that the student has mastered the use of the word. But this does not entail that there need be *criteria* for "X learned the word *w*." All that is required is that we must sometimes have good reasons for saying that the word has been mastered; and this condition is satisfied when, for example, the simplest and most plausible explanation available of the verbal behavior of the student is that he has learned the use of the word.

University of California, Berkeley
and
Center for Advanced Study in the Behavioral Sciences,
Stanford University

²¹ Among the many psychological studies relevant to this point, the following are of special importance: F. Bartlett, *Remembering, A Study in Experimental and Social Psychology* (Cambridge, 1932); J. Piaget, *The Child's Conception of the World* (London, 1928); J. Brunner, "On Perceptual Readiness," reprinted in *Readings in Perception*, ed. M. Wertheimer and D. Beardsley (Princeton, 1958), pp. 686-729.

IV. A VINDICATION OF SCIENTIFIC INDUCTIVE PRACTICES

BRIAN ELLIS

THIS essay is divided into four parts. The first is a discussion of the attempts of Reichenbach and Salmon to find a pragmatic justification of induction. The presuppositions of this approach are set out, and it is shown that the rational preferability of using certain inductive rules might be demonstrated in this way only if it is assumed that the things with which we have to deal are *theoretically isolated*, i.e., that we can say whatever we like about them, without it affecting, in any way, our understanding of the rest of nature.

It is also shown (in Section II) that where the rational preferability of using certain inductive rules might be demonstrated (i.e., where our subject matter is theoretically isolated), the knowledge of which rules we are to use does not help us *at all* to decide what predictions we should make. Therefore, no vindication of induction which presupposes the theoretical isolation of the subject matter of our inductive arguments can possibly succeed. On the contrary, it appears that *theoretical involvement* (as opposed to theoretical isolation) is a necessary condition for the possibility of rational non-demonstrative argument.

That being the case, the question of *what kind* of theoretical involvement is rationally preferable arises. It is shown (in Section III) that if a given sequence of events (e.g., a sequence of results of observations) is, *as it stands*, in conformity with currently accepted scientific theory, we are faced with two, and only two, rational alternatives for projecting this sequence (into the future):

- (a) to project it in such a way that the projected sequence is in conformity with currently accepted scientific theory, or
- (b) to devise an alternative theoretical framework that accounts for the original sequence of events, and for everything else that could be accounted for on the original theory, and to project the sequence in

conformity with this alternative theoretical framework.

But this is now an accurate account of scientific inductive procedures. The alternative (a) describes the ordinary case of using accepted theories to make predictions, and (b) describes the procedure of a Galileo or a Newton in making a theoretical breakthrough. Hence it seems that we have found a vindication of scientific inductive practices.

Some attempt to clarify this proposal for vindicating scientific inductive practices is made in Section IV.

SECTION I

The Scope of the Pragmatic Justifications of Induction

The problem of induction has two parts. First, there is the problem of justifying scientific *methods* of predicting the unknown. Second, there is the problem of justifying scientifically based *beliefs* about the unknown, and showing that, for the most part, at least, they are true beliefs. The first part has come to be known as the problem of *vindicating* induction, and the second that of *validating* induction.¹

H. Reichenbach's pragmatic justification of induction² was an attempt at vindication only. It was not, nor was it claimed to be, anything other than a justification for using certain principles of inference—principles upon which, he considered, all scientific methods of predicting the unknown were based. It was not a proof that all, most, or even some conclusions drawn using these principles must be true.

Very briefly, the argument attempted to show:

- (a) that there is a certain class of inductive rules (convergent rules) persistent use of which must

¹ For a discussion of this distinction, see J. J. Katz, *The Problem of Induction and its Solution* (Chicago, University of Chicago Press, 1962), ch. II.

² H. Reichenbach, *Experience and Prediction* (Chicago, University of Chicago Press, 1938), sec. 42; *Theory of Probability* (Berkeley, University of California Press, 1949), sec. 87.

- eventually yield knowledge of probabilities, provided only that such knowledge is possible, and,
(b) that there are no other rules that share this property.

Then, since it is more rational to adopt a procedure that *must* eventually succeed in achieving our agreed aims, if our aims are capable of achievement, than to adopt one that need not succeed in any circumstances, the persistent use of convergent rules is vindicated.

This argument has been seen to fall short on two accounts. First, it did not succeed in selecting a *particular* inductive rule. Indeed, the class of inductive rules vindicated by Reichenbach's argument is so wide that every hypothesis concerning probabilities is consistent with some choice of convergent rule.³

Second, as probability is explicated by Reichenbach, the argument failed to show that knowledge of probabilities is worth having. For, crudely, if "probability" is understood to mean the practical limit of a sequence of relative frequency estimates (long-run relative frequency), it seems that knowledge of such limits can be of no real interest, unless we can be assured that these limits are *now being approached*. But then, it seems, in order to gain such assurance, we must already have solved the problem of *validating* induction. For this assurance concerns our beliefs, not merely our methods.

We may put this objection by saying that Reichenbach's vindication of induction fails to deal with the problem of the *short run* which is, necessarily, our only real concern. The question of what may happen in the indefinite future may have a certain speculative and dreamy charm. But the only question of real importance is what is going to happen tomorrow, next year, or in the next couple of generations.

However, W. Salmon seems to have shown how both of these objections can be met. The class of convergent inductive rules may be narrowed down to the so-called *straight rule* if only it is demanded that the various probability estimations yielded by the consistent use of a convergent rule (at any

particular time) must be mutually compatible, whatever language is used to describe observed events.⁴ The problem of the short run may then be overcome by another pragmatic justification.⁵ Using an argument akin to Reichenbach's, Salmon claims to have shown that the only permissible short-run rule is the straight rule: Estimate the relative frequency in any finite initial segment of a relative frequency series to be as near as possible to the (assumed known) long-run relative frequency.⁶

Now, we shall not question the validity of Reichenbach's and Salmon's arguments here. But we must consider what they legitimately claim to have done. They have sought a single principle of inductive inference that possesses certain undoubtedly desirable formal characteristics, and they have shown that the only principle having these characteristics is that of *straight induction*. They have not shown that all or even most scientific methods of predicting the unknown are based upon the acceptance of this or any other single inductive principle. Nor have they shown that scientific methods of predicting the unknown (i.e., scientific inductive methods) must yield mutually consistent predictions, whatever language is used to describe events. (Indeed, it is evident that scientific inductive methods do not have this highly desirable property. For, if they did, scientific research would be better done by computers.) Also, and this is most important, they have not shown that if the straight inductive principle is accepted, there is no question as to how, or in what circumstances, it should be applied. Yet this is far from evident.

Consider, for example, the simple case of tossing a coin, which we have examined carefully and found it to be *homogeneous* in substance and *symmetrical* in figure. If 550 heads appeared in a sequence of 1,000 throws, then the straight inductive rule, ordinarily applied, would instruct us to predict that the long-run relative frequency of heads would be 55/100. But, even if no one had ever tossed a coin before, such a projection of the

³ Reichenbach himself made this point, but although he attempted to resolve the difficulty by means of his concept of *descriptive simplicity*, he never really succeeded. For a full discussion, see W. Salmon, "The Predictive Inference," *The Philosophy of Science*, vol. 24 (1957).

⁴ This condition, it will be noted, includes Salmon's *linguistic invariance* as well as his *normalizing* conditions. See: W. Salmon, "Vindication of Induction," *Current Issues in the Philosophy of Science*, ed. H. Feigl and G. Maxwell (Holt, Reinhart and Winston, New York, 1961), pp. 245-264.

⁵ W. Salmon, "The Short Run," *Philosophy of Science*, vol. 22 (1955).

⁶ In fact, Salmon has not shown this. He has succeeded only in restricting the class of permissible short-run rules to that of the convergent ones. However, his argument may be extended, on analogy with his own extension of Reichenbach's, and the class of permissible rules may thus be narrowed down to the straight ones.

sequence would be absurd. For, in the first place, the occurrence of the *observed* sequence is in conformity with our present concepts of physical symmetry and mechanical causation (in the sense that it is the kind of result that is to be expected by anyone who shares our conceptual and theoretical framework). In the second place, the projected limit of 55/100 is inconsistent with this framework. For, if the *limit* 55/100 were seriously entertained, some modification of currently accepted views of physical symmetry or mechanical causation must be envisaged. For example, we might have to suppose that facial markings affect mechanical behavior. But this supposition would certainly have far-reaching theoretical repercussions. Therefore, we cannot accept the projected limit of 55/100, unless we are prepared to make radical revisions to our theoretical and conceptual framework.

It seems clear, therefore, that some restriction must be placed upon the *range of applicability* of straight inductive rules. For the above example shows that the use of straight inductive rules is not always the most rational policy.

It may be thought that in this case the use of straight induction is illegitimate, only because it fails to take into account the many other straight inductive arguments that might be brought to bear on the same issue. In other words, it might be suggested that, if due account were taken of these other (concatenated) arguments, a more acceptable conclusion would be reached.

But this reply is demonstrably inadequate, as is shown by the following argument:

- (1) Scientific theories have a legitimate role in determining what probability judgments we should make (i.e., it is rational to be guided by currently accepted scientific theories in making predictions).

(Hypothesis)

- (2) There are no *determinative* rules for scientific theory construction. That is, there are no rules which determine *uniquely* what theory should be constructed on the basis of what *evidence* (where "evidence" is understood to mean "known facts about particulars").

(Hypothesis)

- (3) An inductive rule is any determinative rule for making probability judgments *solely* on the basis of evidence. (Definition of "inductive rule" which accords with Reichenbach's and Salmon's usage.)
- (4) If all rationally made probability judgments could be justified simply by reference to inductive rules and evidence, then propositions (1) or (2)

must be false. But we are assuming (1) and (2) to be true. Therefore, it cannot be the case that all rationally made probability judgments can be justified solely by reference to inductive rules and evidence.

(*Modus Tollens*)

It follows that no vindication of induction which succeeds only in showing the rational preferability of using certain inductive rules can possibly solve the general problem of vindicating scientific inductive practices. At best, it can only demonstrate the rational preferability of using these rules *where there are no relevant theoretical considerations*. In other words, the most that a pragmatic justification of induction can possibly do is to show the rational preferability of using certain inductive rules to make probability estimates about things that are *theoretically isolated*.

SECTION II

The Vacuousness of the Pragmatic Justifications

Quite apart from the above general restriction on the scope of Reichenbach's and Salmon's vindications of induction, it can be shown that even where there are no relevant theoretical considerations, i.e., where the things we are dealing with are completely isolated theoretically, the mere acceptance of straight inductive rules is utterly powerless to guide us in making predictions. Consequently, where there is no relevant background of scientific theory, every prediction is as good or as bad as every other, and each can be justified on the basis of straight inductive rules. Moreover, it is possible to show that this is so, even *without* introducing such odd predicates as "grue" and "bleen" (as Goodman did in making a similar point).⁷ Hence, our actual *practice* of applying straight inductive rules cannot possibly be justified by any theory of *predicate-entrenchment* (as Goodman hoped).

Suppose that a computer is turning up numbers on a screen, and that the first five numbers to appear are 1, 2, 3, 4, 5, in that order. An observer will naturally expect that if the computer continues to operate, the numbers 6, 7, 8, . . . will subsequently appear on the screen in serial order. But what is the rational basis for this expectation? Let us suppose that our observer has read Reichenbach's and Salmon's arguments and has become convinced that using straight inductive rules is the only rationally justifiable inductive policy. Is he,

⁷ N. Goodman, *Fact, Fiction and Forecast* (The Athlone Press, University of London, 1954), ch. II and IV.

even so, in any position to justify the actual use that he makes of these rules? Of course, he may argue that since in each case $a_n = n$, where a_n is the n th number to appear on the screen, then in future too, $a_n = n$. And, since this accords with the inductive policy which is rationally preferable to all others, this is the most rational prediction that can be made in the circumstances. However, he could equally well argue that since in each case

$$a_n = (n-1)(n-2)(n-3)(n-4)(n-5) + n$$

where a_n is the n th number to appear on the screen, then, in future too,

$$a_n = (n-1)(n-2)(n-3)(n-4)(n-5) + n$$

And, this prediction also accords with straight inductive policy.⁸

Hence, if the prediction $a_6 = 6$ is rationally preferable, on straight inductive grounds, so also is the prediction $a_6 = (5! + 6) = 126$. Indeed, it is not difficult to show that there are infinitely many predictions that can be made, strictly in accordance with straight inductive policies, all mutually incompatible. That being the case, the mere assurance that straight inductive rules are rationally preferable to others, does nothing to vindicate the actual use that we make of these rules, at least, in the kind of case envisaged.

Now, the logical situation is not changed, if in place of 1, 2, 3, 4, 5, the computer throws up the numbers 1, 1, 1, 1, 1. We can argue that since in each case $a_n = 1$, so in general $a_n = 1$. But we can also argue that since in each case

$$a_n = (n-1)(n-2)(n-3)(n-4)(n-5) + 1$$

so, in general,

$$a_n = (n-1)(n-2)(n-3)(n-4)(n-5) + 1.$$

Hence, the prediction $a_6 = 1$ is no better placed than the prediction $a_6 = 121$.

Again, the logical situation is apparently not changed essentially, if in place of the computer sequence, we consider any natural sequence of events. Suppose, for example, that instead of watching numbers thrown onto a screen, we are observing the variations in brightness of a Cepheid Variable star (the first ever observed) and measuring the time-intervals t_n from maximum to maximum. Tabulating our results, we might obtain the sequence:

$$t_1 = 4 \text{ days}$$

$$t_2 = 4 \text{ days}$$

$$t_3 = 4 \text{ days}$$

$$t_4 = 4 \text{ days}$$

$$t_5 = 4 \text{ days}$$

If we can argue that $t_6 = 4$ days on the basis of straight induction, we can also argue that $t_6 = 124$ days on precisely the same grounds. For while it is true that for each n , ($n = 1$ to $n = 5$), $t_n = 4$ days, it is also true that for each n , ($n = 1$ to $n = 5$),

$$t_n = ((n-1)(n-2)(n-3)(n-4)(n-5) + 4) \text{ days.}$$

It is obvious that similar considerations would apply to any set of quantitative results. The actual length of the sequences considered is quite irrelevant to the structure of the argument. For the plain mathematically demonstrable fact is that any finite initial segment of a sequence can be continued in infinitely many ways, provided only that there is no prior constraint on the complexity of the generating functions that may be employed. Consequently, unless we are prepared to supplement the straight rules of induction by some other principles or rules (such as a principle of simplicity), the knowledge that we are to use only straight rules of induction can give us no guidance at all in making quantitative predictions.

Now what applies to quantitative results also applies to non-quantitative ones. The fact that the first ' n ' members of a given sequence of objects all possess the property ' P ' need not imply (even according to straight inductive canons) that the next member of the sequence a_{n+1} will also possess the property ' P '. For the sequence of results a_1 is P , a_2 is P , a_3 is P , ..., a_n is P can be continued in any way we please, and whatever way we choose, there will be an appropriate rule to generate the extended sequence. It is not even necessary to introduce such odd predicates as "grue" and "bleen" as Goodman did in making a similar point.⁹

All of this can be done entirely within our present language. Thus, the prediction a_{n+1} is Q might be justified in the following way. Suppose that when the first result a_1 is P is obtained, the sequence of results a_2 is P , a_3 is P , ..., a_n is P , a_{n+1} is Q , a_{n+2} is Q , ... is envisaged. Then in each case, up to a_n is P , the results obtained

⁸ The reader can readily see that the complex sequence runs: 1, 2, 3, 4, 5, 126, 727, ...

⁹ *Op. cit.*

would be seen to conform to the envisaged sequence. They would be seen to be alike in this respect. Consequently, we should, entirely in accordance with straight inductive rules, predict that in future too the results obtained will conform to this sequence. Consequently, we should predict that a_{n+1} is Q .

It follows then, quite generally, that the Reichenbach-Salmon vindications of straight induction fail to deal with the problem of vindicating the actual use that is made of straight inductive rules. For these vindications fail to account for the fact that it is one thing to state a rule, and another to state how that rule is to be applied.

In reply to objections made along these lines by S. Barker,¹⁰ Salmon proposed to restrict the class of projectible predicates to *purely ostensive* ones.¹¹ A purely ostensive predicate P being any which can be defined ostensively, whose positive and negative instances for ostensive definition can be indicated non-verbally, and such that things possessing P resemble each other in some *observable* respect. Thus, "blue" and "green" are purely ostensive predicates (according to this definition), while "grue" and "bleen" are not. Hence, the Goodman paradox is avoided. But this reply is hardly adequate. In the first place, it puts a severe limitation on the *scope* of the pragmatic justifications of induction. There are many predicates, e.g., "is ten years old," "has a momentum of 10 gm.cm./sec.," "is electromagnetic," which are clearly not purely ostensive but which, we feel, are projectible. Therefore, while the restriction to purely ostensive predicates succeeds in eliminating Goodman's "pathological" predicates, it also eliminates some perfectly normal ones.

Second, it is not evident that such predicates as "is a term in the sequence generated by $f(n)$ " cannot be taught purely ostensively, even when f is a fairly complex function. At least, if simple sequences can be defined purely ostensively, it seems to be only a *contingent* matter, depending on how *intelligent* we are, whether or not complex sequences can also be. Therefore, according to Salmon's defence, if we are not very intelligent, then only simple applications of inductive rules are justified (since, to us, only simple predicates are purely ostensive). If we are *very* intelligent, then quite complex applications of inductive rules are justified (since, now, quite complex predicates

may be learned purely ostensively). If we are super beings with super intelligence, then, presumably, knowledge of inductive rules is utterly useless. For, to such beings, simple and extremely complex applications of inductive rules are equally justified. Salmon's reply to Goodman and Barker therefore works only on the assumption that we are not very intelligent.

Third, the requirement that projectible predicates be purely ostensive seems in one sense to be *ad hoc*. Even if, in fact, it were our practice never to use other than purely ostensive predicates in applying inductive rules, then *what would be the vindication of this practice?* Therefore, even if the first and second of these objections can be met, the problem of vindicating our practice of *applying* straight inductive rules in the way that we do remains unsolved.

SECTION III

Theoretical Involvement as a Necessary Condition for Rational Non-Demonstrative Argument

Consider the sequence of cases discussed in the last section. It will be noticed that the "odd" applications of straight inductive rules seem to become progressively more irrational. Intuitively, at least, that is what we would say. In the first computer case, for example, we should not be greatly astonished to find the number 126 appear on the screen immediately after the number 5. After all, we know that computers can be programmed to generate extremely complex sequences, and it would not conflict with anything else we think we know about the world to suppose that the computer has indeed been programmed to generate a complex sequence which happens to begin: 1, 2, 3, 4, 5, . . .

Of course, we might argue that there are limits to the complexity of the sequences which even computers are able to generate. Moreover, even if there were no such limits, extremely complex sequences like those described are neither mathematically nor physically interesting and, hence, are unlikely to be programmed for any purpose other than trickery. In other words, we might attempt to justify our expectation that the sequence, 1, 2, 3, 4, 5, will be continued 6, 7, 8, . . . in terms of the psychology and interests of computer programmers. Even so, it would lead to

¹⁰ S. Barker, "Comments on Salmon's 'Vindication of Induction'," *Current Issues in Philosophy of Science*, ed. Feigl and Maxwell (Holt, Rinehart and Winston, New York, 1961), pp. 257-260.

¹¹ W. Salmon, "On Vindicating Induction," *Philosophy of Science*, vol. 30 (1963).

no basic conflict with our understanding of the world to suppose that the sequence, 1, 2, 3, 4, 5, would indeed continue, 126, 727, . . . and so on.

In this sense, then, the computer sequences are *theoretically isolated*. If it were not for such incidental information as we may possess concerning the structure of computers, and the interests of computer programmers, we should, indeed, have no grounds whatever for preferring one continuation of the sequence to any other.

The theoretical isolation of the computer sequences (which admittedly is imperfect) stands in striking contrast to the theoretical involvement of what I have called natural sequences. For all that the use of straight inductive rules can tell us, the sequence of time-interval measurements (between maxima in star brightness variations) might be expected to continue in any way at all. But if the sequence were in fact to continue in some "irregular" fashion (that is, in a way that is radically different from what we should ordinarily expect), this would at once pose an immense theoretical problem.

So long as the sequence is "regular," we can see that a detailed quantitative explanation might be forthcoming in terms of accepted laws and theories. Perhaps, we do not yet have a completely satisfactory explanation of the variations in brightness of Cepheid Variables. Nevertheless, the law of brightness variation that we suppose applies to these stars is such that we are able to imagine several possible explanations conforming to accepted laws and theories. But if the law of brightness variation were highly complex, it would be difficult to suggest *any* hypotheses compatible with accepted laws and theories which could possibly lead to the generation of such a sequence. Hence, we must contemplate the possibility that extraordinary laws, quite unlike any others known to us, are involved in the detailed explanation. Either that, or else that our theoretical picture of stellar constitution is grossly inaccurate. We might, for example, contemplate the possibility that stars have a structure comparable in complexity and organization to a computer. But in any case we must consider making radical revisions or additions to the structure of our physics.

Now, if a sequence of events, physically so isolated as the brightness variations of a distant star, is so heavily involved theoretically, the contrast between theoretical isolation and theoretical involvement is even more striking when we turn to consider more everyday sequences. The dis-

covery of a substance which, chemically and physically, possessed all of the known characteristics of lead but which, spectrographically, was utterly different from lead would have far-reaching though unforeseeable theoretical repercussions. If green things everywhere turned blue, and blue things green, we should be completely at a loss to account for such happenings. (That is, if things really were discovered to be grue and bleen, the ramifications of this discovery would extend throughout physics, chemistry, and physiology.) If the length of the day, as measured on ordinary clocks, suddenly began to show enormous variations, this too would have shattering theoretical consequences.

Clearly, if sufficiently many such devastating things were actually to occur, the whole edifice of scientific achievement would come tumbling to the ground, and, for a time at least, anyone's guess as to what would then happen would be as good as anyone else's. Science can take a few shocks; but after sufficiently many, they would cease even to be shocks. For the scientific structure, against which they appeared as shocks, would cease to exist.

In any world where such things have actually occurred in sufficient number, rational argument concerning future contingencies would become impossible. Sequences would occur, but they would have lost their theoretical involvement. Consequently, we should all be in the position of the man watching the computer. Indeed, our position would be somewhat worse than his, since we should have no knowledge of the structure of the computer, or of the interests of computer operators. In a world where every sequence is theoretically isolated from every other, every projection into the future is as sound as every other.

We may, therefore, conclude that theoretical involvement is a necessary condition for rational non-demonstrative argument. This is really the lesson to be learned from the Goodman paradoxes. The paradoxes arise only because we vacillate in our way of regarding such terms as "emeralds" and "green." Thus, on the one hand, we are invited (by Goodman in presenting his example) to agree that all evidence that emeralds are green is, at the same time, evidence that all emeralds are grue. (To comply, we must take "evidence" to mean "straight inductive evidence," and we must disregard the theoretical commitments made by the use of the terms "emeralds" and "grue.") On the other hand, it is suggested that the con-

clusion that, after a time t all emeralds will be blue, is absurd. (But it is so only if "emeralds" and "blue" are *not* taken to be meaningless, theoretically uncommitted terms, but to refer to a certain kind of *stone* and a certain *color* respectively.)

Now, since no rational inductive policy can be self-defeating, it follows that the only permissible inductive policies are those whose use would not destroy the theoretical involvement of our concepts. For, without theoretical involvement, rational non-demonstrative argument is impossible. We may therefore argue:

- (a) It is irrational to reject any scientific theory solely on the basis of future or otherwise unobserved possibilities. (For, if this were rational, all scientific theories could be rejected without further ado, and then, as we have seen, rational argument concerning future contingencies would be impossible.)
- (b) Therefore, if any given sequence of events already conforms to currently accepted scientific theory, then it is irrational for anyone to prefer any projection of the sequence that does not conform to the currently accepted theory unless they have an alternative and so far equally satisfactory theoretical framework.
(For to do so is to take up a position requiring the rejection or revision of scientific theories solely on the basis of future or otherwise unobserved contingencies. If this were a rational procedure, then, as before, rational argument concerning future contingencies would be impossible.)

SECTION IV

Discussion

The vindication of scientific inductive practices proposed at the end of the last section is open to a number of objections and is clouded by some ambiguities. First, the use of the term "conformity" needs to be explained. When we say that something conforms to our current conceptual and theoretical framework, we mean that it is *the kind of thing that is to be expected* by anyone who shares this framework. Thus, if a coin is tossed, we should expect it to land heads or tails; both results are in conformity with our conceptual and theoretical framework. But if the coin should cease to be a coin and be transformed before our eyes into a butterfly, then this result would not conform to this framework. If people really believed that such a thing had happened, their faith in scientific achievement would be shattered.

There is, however, no hard line to be drawn between conformity and disconformity. The examples taken are extreme ones. Between these extremes there are many intermediate and borderline cases. A disconforming sequence is any which we should judge to be either *physically impossible* or *physically incongruous*; a physically impossible sequence being any which conflicts *directly* with some accepted law or theory, and a physically incongruous one being any which we should judge to be inexplicable on the basis of accepted laws and theories. The "odd" projection of the sequence of brightness variations discussed in Section II is an example of a physically incongruous sequence.

If a sequence of events neither conforms nor disconforms to currently accepted theory, then it is *theoretically isolated*. The best examples of theoretically isolated sequences that I have been able to construct are the computer sequences discussed in Section II.

Now, not only is there no sharp distinction between conformity and disconformity but different scientists may disagree about what conforms to what. Thus, certain results which Count Rumford thought were utterly fatal to the Caloric Theory were not in fact taken to be so by his contemporaries. And it is hard to put this down either to ignorance or dishonesty. Nevertheless, it is in this sense that we say that if any sequence of observed events is already in conformity with accepted scientific theory, it is irrational to project this sequence in any way that will bring it into disconformity with this theory. If, in fact, the sequence is so short, or the theoretical involvement so tenuous that several conformable projections are possible, then the choice between them is arbitrary. And, in the limit, where there is no theoretical involvement at all, no one projection is to be preferred to any other. Where there is a disagreement about conformity, then we also have a disagreement about who is being rational.

Now it may be objected that if, as I have argued, the possibility of making rational decisions about the future depends upon the existence of a theoretical superstructure then, in the pre-theoretical stages of human development, such decisions would have been impossible. It is doubtful whether there ever was such a stage of human development, at least while men were still recognizable as men. But if there were such a stage, I see no reason why this conclusion should not be accepted. There was, after all, a time when the intellectual leaders among men were seriously concerned whether the

sun would rise the following day and they propitiated their gods to ensure that it did so.

It may also be objected that my argument is incomplete since I have justified the use that is actually made of inductive rules by reference to the currently accepted theoretical superstructure, but I have not explained why this superstructure should, in general, be accepted in preference to any other. This is quite true. And if this objection could not be met, it would prove a serious objection to my proposed vindication of scientific inductive practices.

Consider first the case of a man who has a real alternative theoretical framework to offer us, who can show that presently observed sequences of events are as much, or more, in conformity with it than with our present theoretical scheme, and who is able to make predictions on the basis of his alternative scheme different from those that we should make, but which are subsequently borne out by experience. Such a man has made a theoretical break-through in science. If it involved sufficient revision of accepted theory, then it might be compared with the seventeenth century break-through in astronomy and dynamics. Now, obviously, *he* is not obliged to accept the theoretical superstructure that has so far been evolved. He has something *demonstrably* better to offer us.

I want to say that such a man has succeeded in changing our standards of rationality. Consider the theoretical break-through which led to the overthrow of the Caloric Theory of heat and its replacement by the Kinetic Theory. If, today, a man were to use the Caloric Theory of Heat to make predictions in situations where there are heat-work exchanges, then he must be either ignorant or irrational. If he is unaware of the nature of the events which led to the replacement of the Caloric Theory, then he is ignorant. If he is aware of these events, and still persists in using the Caloric Theory (I do not mean a *revised* version of the Caloric Theory), then he is simply irrational. For he would, knowingly, be making predictions which, if true, must lead to the rejection of currently accepted and, so far, satisfactory theories of heat and work. Hence, by implication, he must be prepared to reject such theories solely on the basis of future (unrealized) possibilities.

Next, consider the case of the man who has no real alternative theoretical framework to offer us but who, nevertheless, refuses to accept our current standards of rationality, even though he is thoroughly informed about all phases of the history

of science. Such a man cannot be rational. If he accepts no theoretical framework at all, then to him, all events are theoretically isolated, and he is in no position to make rational predictions. He is genuinely in the position that Hume imagined all men to be. If he accepts any other theoretical framework, e.g., that of 200 years ago, then, as we have seen, his method of reasoning is such that he is prepared to reject theories that have so far proved to be entirely satisfactory solely on the basis of future possibilities.

Since this is an important point, it deserves emphasis. Consider the case of two men, *A* and *B*, arguing about the likely outcome of an experiment in which heat-work exchanges are involved. *A* is a Caloricist who views the production of work in a heat engine as due to the falling of heat from a higher to a lower (temperature) level, much as work is produced in a water mill by the falling of water from a higher to a lower level. *B* is a Kineticist who views the production of work as due to the conversion of heat into work. Let us suppose that their predictions differ. What reason do we have for preferring *B*'s prediction to *A*'s? We may admit at once that anything might happen. But the following points must be granted:

- (i) It would be irrational for me to accept or reject any theory solely on the basis of what might happen.
- (ii) *A*'s Caloric Theory cannot be accepted on the basis of what has happened. (I am assuming that *A*'s theory is the historically accepted theory of the 1830's, not a radically revised theory.)
- (iii) *B*'s Kinetic Theory *can* be accepted on the basis of what *has* happened.
- (iv) There is, as yet, no other theory that can be accepted on the basis of what has happened.

From these four propositions it follows at once that we *must* accept *B*'s reasoning if we are to behave rationally. If we accept *A*'s reasoning then we are committed to accepting *A*'s theory and rejecting *B*'s theory entirely on the basis of future possibilities. If we reject *B*'s reasoning but do not accept *A*'s, then again we are in the position of rejecting a theory simply because of the logical possibility that anything might happen. If we do not accept any reasoning, then obviously the possibility of rational argument is precluded. Therefore, if we are rational beings we have no alternative but to accept *B*'s reasoning until some alternative and so far equally satisfactory theory is proposed.

The difficulty remains, however, of saying what counts as an alternative and equally satisfactory theory. If there are no grounds for rejecting theories other than:

- (a) inconsistency,
- (b) conflict with experiment or observation, or
- (c) incompatibility or incongruity with other already accepted laws or theories,

then the construction of an alternative theory might seem to be an easy matter. Could we not follow the method of *goropising* described by L. S. Feuer¹²—viz., make any hypothesis we like, and then, for each new piece of contrary evidence that comes to hand, make a special *ad hoc* hypothesis to account for it? Could we not construct “theories without analogies” after the fashion of N. R. Campbell’s famous example?¹³ Yet, if the construction of alternative theories is such an easy matter, then

almost any conclusion about the unknown that we may care to draw could be justified by reference to some (*ad hoc*) theoretical superstructure. And hence, our proposed vindication of scientific inductive practices is seriously incomplete.

It is not evident that “goropising” is in fact such an easy method of constructing a “satisfactory” alternative theory. For typically, “goropising” leads in a few steps to hypotheses that are *incompatible* or *incongruous* with other already accepted laws or theories. Nevertheless, a residual problem remains: we must either show that the grounds (a) to (c) are sufficiently restrictive to make the construction of alternative and so far “equally satisfactory” theories generally a matter of great difficulty; or else, we must say what principles in addition to (a), (b), and (c) are involved in the selection of theories; and then, we must justify using these principles in the way that we do.

University of Melbourne

¹² L. S. Feuer, “The Principle of Simplicity,” *Philosophy of Science*, vol. 24 (1957).

¹³ N. R. Campbell, *Foundations of Science* (New York, Dover, 1957), p. 123.

V. PROPOSITIONS AS ANSWERS

J. E. LLEWELYN

MANY writers have recently considered it necessary to issue the reminder that a formal system does not stand condemned simply because it does not fit the contours of ordinary discourse, since the mapping of these contours may not have been among the purposes for which that system was constructed. In particular, dispute as to whether failure of reference implies that on a certain occasion a sentence would express a false proposition or no proposition at all may reflect nothing more serious than a difference between a decision to use the words "proposition" or "false" in one way and a decision to use them in another.¹ It is a dispute closely related to this one that is examined in the present article. It too hinges upon a criticism of what is asserted or assumed about truth-value by Russell and others. This criticism is implicit in Collingwood's statement (*S*₁) that what is ordinarily meant when a proposition is said to be true is that it is the "right" answer to a question which arises.² Since this is about "what is ordinarily meant" Russell could plead the same defense as he used against Strawson's remarks on referring. That this is so is one of the things that my discussion will illustrate. This discussion has three stages. I shall first argue that *S*₁ must be rejected. After mentioning a more inclusive thesis of Collingwood's (*S*₂) which is *a fortiori* unacceptable the article concludes with a consideration of an amended version of *S*₂ (*S*₃).

I

On one natural interpretation of *S*₁, there is evidence which counts strongly against it. A proposition is sometimes said to be true when it is the answer to a question which is silly because it is irrelevant and which, therefore, in one sense of the phrase, does not arise. In order to see that this

is so, imagine that a group of archaeologists is investigating whether the Middle Minoan palace at Knossos was occupied by the pre-Greek dynasts. If one of the group asks out of the blue whether kangaroos are born inside or outside of the mother's pouch his question might well provoke the retort "Outside. But what has that to do with it?" If to this the questioner says "Nothing at all," and if he is sincere in saying this, it would mean that in one sense his question did not arise either for his colleagues or himself. It was not an intelligent thing to ask. It was a pointless enquiry. Nevertheless it would be wrong to deny that "Outside" was a true answer. We should so describe it in spite of the fact that the question it answered was not "what we ordinarily call a sensible or intelligent question." Hence, in order that *S*₁ may be reconciled with this fact the conditions under which a question arises must be defined more narrowly than Collingwood has defined them here. He himself does this elsewhere.³

What is the "right" answer to a question? Collingwood tells us that "Cases are quite common in which the 'right' answer to a question is 'false'." This is so when the false answer serves to bring out a false presupposition of the question. The example he gives is the answer which Polemarchus made to Socrates' catch question whether he would prefer to play draughts with a just man or with someone who knows how to play draughts. Polemarchus replied that he would prefer to play with someone who knows how to play. This reply enabled Socrates to make the point that Polemarchus is presupposing that justice and the capacity to play draughts are both special skills. It is not obvious whether Collingwood intends to imply that the other of the two answers that a question like this invites us to choose between is true and wrong, false and wrong, true and right,

¹ See, for example, W. V. O. Quine, "Mr. Strawson on Logical Theory," *Mind*, vol. 62 (1953), pp. 443-446, and Bertrand Russell, "Mr. Strawson on Referring," *Mind*, vol. 66 (1957), pp. 387-389. Cf. Gilbert Ryle, *Dilemmas* (Cambridge, Cambridge University Press, 1954), pp. 111-129.

² R. G. Collingwood, *Autobiography* (London, Oxford University Press, 1939). References are to this where not stated otherwise.

³ *Essay on Metaphysics* (Oxford, Clarendon Press, 1940), pp. 25-33 and 38-48.

or that it is false and right. If *S1* is taken in conjunction with the explanation offered in the *Essay on Metaphysics* of what it is for a question not to arise, it follows that Socrates' question does not arise and hence that no answer to it could be true. One of the obstacles in the way of deciding where Collingwood stands on this matter is his practice throughout the chapter entitled "Question and Answer" of sometimes putting the word "false" in shudder-quotes and sometimes not. But since on the same page as that on which he describes the answer made by Polemarchus as "false" (in quotes) he also uses the word within the quotes of a proposition which would be an untrue answer to a question which has no false presupposition, it would appear to be Collingwood's intention to describe the answer made by Polemarchus as untrue, rather than as, for instance, neither true nor false. Whether this is so or not, his statement about what is ordinarily meant when a proposition is said to be true falls short of providing sufficient conditions for a proposition's being true. A false proposition could fulfill the conditions he mentions.

It might be objected that neither of the two points just made has any bearing on what Collingwood says about true propositions since both presuppose that it is propositions that are true, whereas he says that it is only to the question-and-answer complex of which a proposition is a part that the word "true" can properly apply. There is, however, enough evidence to show that he was not here propounding the odd view that "true" applies only to such combinations as "Is this a leap year?—Yes, this is a leap year." When he says that "true" applies to a question-and-answer complex he means that it applies to a proposition only in so far as it is regarded as an answer to a particular question. Meaning, agreement, contradiction, truth and falsity, he states, belong only to propositions as the answers to strictly correlative questions; they belong to propositions, but not to propositions "in their own right" (p. 33). Let us call this statement *S2*.

S2 is plainly incorrect if it is understood as a claim that the semiotic predicates it mentions cannot belong to a proposition unless that proposition is an answer to a question that has in fact been raised. This is apparently what Collingwood is claiming when he writes that "... you cannot find out what a man means by simply studying his

spoken or written statements, even though he has spoken or written with perfect command of language and perfectly truthful intention. In order to find out his meaning you must also know what the question was (*a question in his own mind, and presumed by him to be in yours*) to which the thing he has said or written was meant as an answer" (p. 31, my italics). But an omniscient being would know propositions to be true without his needing to ask questions of himself or anyone else. And many an elementary analytic proposition is known to be true by ordinary mortals without being answers to questions raised by those who assert them. Similarly for many synthetic propositions whose truth is taken for granted (not called into question) and for such affirmations as "I am over twenty-one" uttered as a rejoinder to "You aren't over twenty-one." And if trite remarks like those we make about the weather are always answers to questions that are in the speaker's mind, it is unduly pessimistic to allege that these questions are presumed to be also in the minds of the people addressed.

Perhaps it was Collingwood's experience as an archaeologist that misled him into the belief that in asserting a proposition we are always reporting a discovery. He is impressed with the need for the researcher to "get ahead with the process of questioning and answering" (p. 37), to state the proposition or make the supposition that has "logical efficacy," i.e., that "causes a certain question to arise."⁴ Consequently he thinks of knowing, and not just researching, as an activity (p. 30). This is a mistake that I do not intend to discuss. Instead I shall bring this purely critical part of my enquiry to a close by drawing attention to the sympathetic interpretation of Collingwood's account given by A. D. Ritchie⁵, who suggests that Collingwood may be recommending that we regard hypotheses as yes-or-no questions, or contending that they are such questions. I shall refer again to Ritchie's comments. Here I limit myself to mentioning one objection that would have to be met by anyone arguing for this analysis of hypotheses: although we verify or falsify hypotheses, we do not verify or falsify a question—we answer it. Even if we say that a hypothesis is a supposal and that "supposing, unless it is mere day dreaming, means asking the question 'Is it true?' and trying to find an answer," it does not follow that the supposal and the supposing are the same. We can

⁴ *Ibid.*, p. 27.

⁵ "The Logic of Question and Answer," *Mind*, vol. 52 (1943), pp. 24–38.

ask "Is what true?" and the answer to this question would cite the hypothesis that is up for testing, and that it is possible to ask these questions of it shows that we are not dealing with the kind of hypothesis that is postulated or "supposed in the sense of being taken as if true and therefore not needing any verification."⁶

II

Would it be correct to say (*S*₃) that the semiotic predicates listed in *S*₂ cannot be known to apply truthfully to given propositions unless it is known to which strictly correlative question each proposition *could be* an answer?

There is a sense in which "My name is Richard" could be an answer both to the question "Whose name is Richard?" and to the question "Is your name Richard?"; for if *A* asked *B* "Whose name is Richard?" and *B* replied "My name is Richard," there would be no need for *A* to ask *B* "Is your name Richard?" But let us understand "could" in *S*₃ in the sense in which "My name is Richard" could be an answer to "Whose name is Richard?" and could not be an answer to "Is your name Richard?", the sense which explains why we might suppose that *B* had misheard if he said "My name is Richard" in response to *A*'s "Is your name Richard?"

Consider, first, agreement and contradiction. Collingwood alleges that "it is impossible to say of a man, 'I do not know what the question is which he is trying to answer, but I can see that he is contradicting himself'" (p. 33). This, it seems to me, is not so. Provided I know that there are or could be answers to the same question I can know whether two propositions are contradictories without knowing what that question is. Under these conditions this would be so of "No Italians have blue eyes" and "Some Italians have blue eyes." And if two propositions are describable as being in agreement when they are not mutually contradictory, whether or not they are or could be answers to the same question, we can know if they are in agreement even when we are ignorant of the question or questions to which each could be an answer. We can know this, for example, of the subcontraries "Some Italians do have blue eyes" and "Some Italians don't have blue eyes" in spite of a mistaken belief that the only question each of these propositions could answer is "Do any Italians have blue eyes?"

⁶ *Ibid.*, p. 29.

Take next the case of meaning. I assume that "Yes" and "No" state propositions only if they function as abbreviations. If they have cognitive meaning it cannot be known what that meaning is unless it is known what they abbreviate. So let us confine ourselves to unabbreviated forms like "The time is seven o'clock." The meaning of this sentence can be known without knowing whether on a particular occasion of its use it is meant as an answer to the question "Is the time seven o'clock?" or to the question "What is the time?" And here what concerns us is only the meaning of words. If we know the meaning of the words in "The time is seven o'clock" we already know the meaning of the words in "Is the time seven o'clock?" We have seen, however, that Collingwood is not talking about the meaning of sentences or words, but about "what a man means," "his meaning," i.e., what the speaker intends in the sense of what question the thing said was meant to answer. We have reinterpreted this such that knowing what a speaker means or intends is knowing the question to which what he says could be an answer. Thus if what is spoken or written is the sentence "My name is Richard" we do not in the relevant sense know what is meant by this unless we know whether it could be an answer to the question "What is your name?" or, on the other hand, an answer to the question "Whose name is Richard?" To this extent the meaning of a proposition is relative to the question it answers. But this does not imply, as Collingwood supposes, that the truth of the proposition is relative to that question, nor that the truth-value cannot be known if the meaning is not known.

One of the criticisms he makes of the exponents of what he calls propositional logic (under which heading he includes traditional formal logic, idealist logic, and modern symbolic logic) is that they assume one can discover whether a proposition is true without taking into account the question which the proposition is meant to answer. That is, he finds nothing wrong with the belief that propositions can be examined in isolation from the questions they are meant to answer. What is wrong, in his view, is the belief that we can determine the meaning, truth, falsity, etc., of propositions that are thus isolated. Hence, on this view, the proposition that would be propounded in the words "My name is Richard" is the same whether they are spoken in answer to "What is your name?" to "Whose name is Richard?" or to

"Is your name Richard?" But in that case, although we cannot in the relevant sense know what the speaker means when he says "My name is Richard" unless it is known which question this is meant to answer, we can know that if the proposition would be a true answer to "What is your name?" it would also be a true answer to "Whose name is Richard?" And so long as we know who utters the indicative sentence we can verify what he says, notwithstanding our ignorance of what it was precisely that he was intending to say.

III

The point just made still holds if it is maintained that the proposition which would answer one of the three questions is not the same as the proposition which would answer one of the others or the question "Isn't your name Richard?" And this is what we ought to maintain if we intend the word "proposition" to refer to "what is asserted," "a logical, not a linguistic, entity" (p. 31). Collingwood does not consistently carry out his intention and, although he issues a warning against the temptations that grammar holds out to logic (p. 34), he is himself seduced by the idea that the logician's proposition is a double of the indicative sentence of traditional grammar. He supposes that a proposition can first be given and that we can then go on to find out to what question it is or could be an answer. Whereas if a proposition is given we already know what questions it could answer. This is brought out by the fact that if we simply wish to report what someone said and do this in the sentence "He said that *his* name was Richard," it can be inferred that what the original speaker said would have answered neither "What is your name?" nor "Is your name Richard?" but "Whose name is Richard?"⁷ Intonation and stress patterns need to be described or illustrated if we wish to indicate how a sentence was or should be spoken. But, as well as being part of how a sentence is spoken, intonation and stress also indicate what is said.⁸ Other devices help to reduce the risk of mistaking what is said, e.g., the combination of a certain intonation and stress pattern or typographical variation with an impersonal verb phrase and a dependent clause, as in "It's *I* who am

called Richard," and "He said that it was *he* who was called Richard." The introduction of other conventions would make it possible to get the same job done with no changes of intonation or stress at all. Languages are conceivable and probably exist in which a sentence must contain a word or morpheme which prevents that sentence being ambiguous in this manner with respect to the question it could answer.

It is worth observing that if, like Collingwood, we reserve the term "proposition" for what is asserted, the method of identifying propositions by conversion into indirect speech shows that "I was exiled to Elba," "You were exiled to Elba," and "He was exiled to Elba" may all assert the same proposition; and what is said in these different ways could be an answer to the one question askable by means of at least three interrogative sentences in the same language: "Was I exiled to Elba?", "Were you exiled to Elba?", and "Was he exiled to Elba?" On the other hand, this test shows that different propositions are asserted by means of the indicative sentences "He will be exiled to Elba" and "He was exiled to Elba." In general, a proposition may survive the substitution of one non-temporal token-reflexive expression for another, but does not survive the substitution of one temporal token-reflexive expression for another (e.g., the substitution of a differently tensed verb) even when the expressions are equivalent in that they indicate the same point or period of time. Two sentences do not express the same proposition unless they would express the same proposition when spoken at the same time. Thus the rules for tense-sequence permit both (1) "He is exiled to Elba" and (2) "He was exiled to Elba" to be reported in the sentence (3) "He said that he was exiled to Elba." But (1) and (2) cannot be used to say the same thing since when they are spoken at the same time what (1) says can be true while what (2) says is false; and if the verb in (2) were equivalent to the verb in (1), nothing with a truth-value could be said in (2) by uttering it at the same time as (1): "He was exiled to Elba now" doesn't make sense.

What a speaker says, his message, is to be distinguished on the one hand from what he is treating of, the subject or topic of his remark, and on the other from the reference of a word or phrase

⁷ This inference cannot be made where we also wish the report to help resolve what appears to be a confusion about who is who.

⁸ Cf. J. Cook Wilson, *Statement and Inference* (Oxford, Clarendon Press, 1926), pp. 114-126.

that he uses. If p would answer "Where was Napoleon exiled?" we should say that the topic was (that of the place) where Napoleon was exiled. If p would answer "Who was exiled to Elba?" we should say the topic was (that of the person) who was exiled to Elba. It might be thought that topic and reference coincide in the case of statements like "The place where Napoleon was exiled was Elba"; but the place is not the same as the issue, question, subject, or topic of the place. Similarly, the statement that the question of where Napoleon was exiled is not a difficult one has as its topic the question whether the question referred to, viz., "Where was Napoleon exiled?" is a difficult one or not. Now the topic of someone's remark is not part of what he says. But where we are concerned with propositions that may be expressed in a given sequence of words whose meaning remains constant from one use of that sequence to another, then, if the sentence "Napoleon was exiled to Elba," for example, is used to make a remark on the topic of where Napoleon was exiled, what is said will be different from what would have been said if the topic of the remark had been who was exiled to Elba. So what someone says and the topic of his remark are indeed interdependent. This is just another way of saying that a proposition and the question it could answer are interdependent. I have said that Collingwood misconstrues the manner of this interdependence when he supposes that we can be given a proposition and then go on to ask to what question it is or could be an answer. But in addition to misrepresenting the relation of what is said, the message, to the topic, he fails to mark clearly the difference between the topic and the reference. For he argues that the appearance of contradiction in "The world is both one and many" can be removed once we know that two questions are being answered here: a question about what is in the world and a question about the world as a totality or container. However, why take this round-about way of bringing out the ambiguity in the phrase "the world"? Why not go directly to this phrase as it occurs in the indicative sentence and point out that it is being used with ambiguous reference and hence that two propositions are being asserted, not one? For ambiguity of reference can infect questions no less than propositions.⁹

According to Collingwood's analysis of what is

asserted, "You cannot tell what a proposition means unless you know what question it is meant to answer" (p. 33). According to my analysis a proposition has built into it an unambiguous indication of the question it could answer—though the question it unambiguously indicates may be ambiguous, as I have mentioned in the immediately preceding paragraph. If this is acceptable to the practitioners of what Collingwood calls propositional logic, his criticism is misplaced. But it may be argued: "This cannot be what the exponents of this logic mean by 'proposition.' The prepositions of which logic treats are not what is asserted, for no one asserts them. It would be absurd to ask what was meant by such a proposition, i.e., to ask to what specific question it was intended as an answer, since it was not intended as an answer to any question at all." With this I should agree. If it is correct to say that formal logic treats not only of propositional schemata or propositional functions but also of propositions, it needs to be recognized that these propositions are stand-in or dummy ones. The sentences of a syllogism in traditional formal logic, for example, are not used there to make assertions, any more than Olivier is asserting something when he recites the lines that Shakespeare wrote. And the syllogisms that appear in the logical treatise are not arguments but patterns or forms that arguments can follow or fill. Hence, if we mean by "proposition" what Collingwood means by it or what I have meant by it for the purposes of engaging with him in this discussion, if "the logical entity" represented by the sentence "He was exiled to Elba" is in this sense a proposition, no proposition whatsoever is propounded in this sentence where it occurs in a syllogism in a book on traditional formal logic. We should have to say that in this context the sentence represents what could be asserted, what we might, accepting Collingwood's own suggestion, call a proponible.¹⁰ Of course the book would also contain propositions in that it would be saying (as well as demonstrating) something about proponibles.

Collingwood's chapter on "Question and Answer" can be interpreted therefore as a complaint that the exponents of "propositional" logic claim to be talking about propositions but are really talking about proponibles. They are scolded for not talking about propositions and their cor-

⁹ Cf. Alan Donagan, *The Later Philosophy of R. G. Collingwood* (Oxford, Clarendon Press, 1962), p. 61.

¹⁰ *Ibid.*, p. 33.

relative questions.¹¹ This is a captious complaint, for in treating of propositions the practitioner of "propositional" logic makes a large number of valuable points about the entailment relations of the propositions in lieu of which they stand. For instance, traditional formal logic and modern symbolic logic are in their different ways able to make the point that in a syllogism of the form: "He was exiled to Elba and Elba is a Mediterranean island, so he was exiled to a Mediterranean island," if the premisses are true then the conclusion is true. This holds good whether the conclusion would answer the question "Was he exiled to a Mediterranean island?" or rather the question "Where was he exiled?" and whether the personal pronoun refers to Napoleon or to Julius Caesar.

IV

In his discussion of what Collingwood says about propositions, Ritchie asks: "Is it impossible for the same proposition to be the answer to different questions? If questions could always be formed so that the answer was 'Yes' or 'No' it would be impossible. But can they?"¹² It follows from my argument that if a proposition is what is said, then even when the proposition is or could be an answer to a question which is not of the yes-or-no kind it could be an answer to only one such question put in the same words—on the assumption being made throughout this argument that the meaning of those words remains constant from one use of them to another. Thus if the proposition "Napoleon was exiled to Elba" could (in the sense distinguished at the beginning of Section II) answer the question "Where was Napoleon exiled?" it could not answer the question "Who was exiled to Elba?" nor could it answer the question "To which Mediterranean island was Napoleon exiled?" For it could be true that Napoleon was exiled to Elba, but false that Elba was a Mediterranean island. Of course the words "Napoleon was exiled to Elba" could be used to answer the question "To which Mediterranean island was Napoleon exiled?"

On the other hand, I am inclined to think that it is not incorrect, but at the most only slightly misleading, to state that "He was exiled to Elba," "You were exiled to Elba," and "Napoleon was exiled to Elba" may all say the same thing.¹³ When all three sentences are used by different people to say something on the topic of where Napoleon was exiled, we can truly report that all three speakers said that Napoleon was exiled to Elba. For this does not imply that any of the three speakers knew the name of the person they were referring to. That the reporter refers to Napoleon by name does not imply that the people reported on referred to Napoleon by name, any more than it is implied that the second speaker used the proper name or pronoun "he" when what the speaker said is reported by stating "He said that he was exiled to Elba" or "He said that he (Napoleon) was exiled to Elba." It is what the speakers said that is being reported, not their words. Indeed, "He said that Napoleon was exiled to Elba" could be a correct report of what someone said even if he did not himself refer to the island by that name or if at the time he spoke there was no island with that name.

Part of the explanation why, under the conditions posited in the first paragraph of this section, "Napoleon was exiled to Elba" and "The Mediterranean island to which Napoleon was exiled was Elba" do not state the same proposition, whereas, under the conditions posited in the second paragraph of this section, "He was exiled to Elba" and "Napoleon was exiled to Elba" do state the same proposition is as follows. In reporting what was said by the person who spoke the words "The Mediterranean island to which Napoleon was exiled was Elba," we use an independent definite descriptive phrase; while if a definite descriptive phrase were used in reporting what was said by the people who spoke the words "He was exiled to Elba" and "Napoleon was exiled to Elba" it would be either a dependent definite descriptive phrase like "the person referred to" in "He said that the person referred to was exiled to Elba," or an independent definite descriptive phrase

¹¹ Cook Wilson is another of the logicians whom Collingwood criticizes for ignoring the question a proposition answers. Compare with this the opinion that Cook Wilson is doing "erotetic" logic, i.e., "is consciously or unconsciously preoccupied with types of questions, rather than with types of statements." See A. and M. Prior, "Erotetic Logic," *Philosophical Review*, vol. 64 (1955), p. 59, and Cook Wilson, *loc. cit.*

¹² Ritchie, *op. cit.*, (n.s.), p. 38. Since "Yes" and "No" could answer both "Is this blue coat yours?" and "Is this blue coat not yours?" Ritchie must be treating this kind of difference as one which does not amount to a difference of enquiry. Nor is it implausible to say that the second differs from the first only rhetorically in that it suggests which answer the enquirer thinks would be true.

¹³ Cf. L. J. Cohen, *The Diversity of Meaning* (London, Methuen, 1962), p. 142.

which the question alone does not entitle us to substitute, e.g., "the man who declared himself Emperor of France in May, 1804."

It seems to me therefore that if the modifications I have recommended are made, Collingwood is substantially correct in saying:

A highly detailed and particularized proposition must be the answer, not to a vague and generalized question, but to a question as detailed and particularized as itself. For example, if my car will not go, I may spend an hour searching for the cause of its failure. If, during this hour, I take out number one plug, lay it on the engine, turn the starting handle, and watch for a spark, my observation "number one plug is all right" is an answer not to the question "Why won't my car go?" but to the question, "Is it because number one plug is not sparking that my car won't go?" Any one of the various experiments I make during the hour will be the finding of an answer to some such detailed and particularized question. The question, "Why won't my car go?" is only a kind of summary of all these taken together. (P. 32)

Not only does "Number one plug is all right" fail to answer "Why won't my car go?"—for we want to know what is wrong with the car, not what is right—but "Number one plug is not sparking" would also fail to answer this question if it succeeds in answering the question "Is number one plug sparking?" What is said in answer to the question "Why won't my car go?" is on the topic of the reason for the car's not going; but the topic of the answer to "Is number one plug sparking?" is the question whether number one plug is sparking. However, if we are to be as precise as Collingwood

urges us to be we should have to disagree with his statement that "Number one plug is all right" is (or could be) an answer to the question "Is the stoppage due to failure in number one plug?" The propositions that (could) answer this question are "No, the stoppage is not due to failure in number one plug" and "Yes, the stoppage is due to failure in number one plug."¹⁴

The passage just quoted suggests that Collingwood might be prepared to argue that variable questions of the form "What . . .?", "Who . . .?", "Where . . .?", etc., are a summary or nest¹⁵ of yes-or-no questions and should be broken down into their yes-or-no components. Compare this passage with the description he gives of his own practice of dividing a yes-or-no question like "Was there a Flavian occupation on this site?" into a series of questions under different heads, e.g., "Are these Flavian sherds and coins mere strays, or were they deposited in the period to which they belong?" (p. 24). If he were ready to advocate this kind of analysis of variable questions, added point would be given to Ritchie's statement that a proposition which could answer a certain yes-or-no question could not answer a different question. The argument I have put forward in this article supports this statement, but it goes further in that it arrives at the conclusion that this holds also for the case of a proposition which could answer a certain variable question. And a corollary of this conclusion is that Collingwood's strictures upon Russell and formal "propositional" logicians in general would be misplaced even supposing, what is false, that he is correct in thinking that these logicians deal directly with what is asserted.

University of Edinburgh

¹⁴ The remark that a proposition ". . . could not be the right answer, to any question which might have been answered otherwise" (pp. 31–32) cannot be read as implying that if a yes-or-no question could be answered affirmatively it could not be answered negatively. Nor can it be taken to imply that if "Yes" is right then "No" cannot be right, i.e., cannot be "the answer which enables us to get ahead with the process of questioning and answering" (p. 37). See section I above.

¹⁵ Collingwood, *op cit.*, p. 38. I use the label "variable question" here to mark a contrast between yes-or-no questions and questions of other kinds. It needs to be noted, however, that a yes-or-no question like "Is the stoppage due to failure in number one plug?" can be reframed as a higher order variable question, e.g., "What is the truth-value of the statement 'The stoppage is due to failure in number one plug?'"

VI. SEEMING TO SEE

CLEMENT DORE

I

TRADITIONALLY, philosophers have believed that there is a certain state—which they have called “having a sense-impression” and “sensing a sense-datum” among other names—the existence of which is logically necessary though not sufficient for the perception of any object. Based on this traditional view of perception is the problem of what proposition, in addition to the proposition that one has a sense-impression (senses a sense-datum, etc.), must be true in order that it be true that one really perceives an object. Phenomenalism and the causal theory of perception may be interpreted as two different attempts to solve this problem.

Many contemporary philosophers take a dim view of sense-impressions and, hence, of the epistemological problem just mentioned. Professor Gilbert Ryle is one such philosopher. He has expressed scepticism about the existence of sense-impressions (“experiences alleged to be basic ingredients in sense-perception” as he calls them) on the grounds that “we have, in fact, no special way of reporting the occurrence of these postulated impressions” and “are, therefore, without the needed marks of our being conscious of such things at all.”¹ Presumably, when Ryle denies that we have a special way of reporting sense-impressions, he has in mind some locution or locutions which are commonly employed in ordinary language as well as in philosophy. (Otherwise, it is difficult to see why Ryle should hold that “sense-impression” itself does not qualify for the job in question.) But now is Ryle right in maintaining that ordinary language contains no devices for enabling us to refer to the sense-impressions of the traditional epistemologists? In this paper, I want to consider an argument on behalf of a negative reply to this question. The argument is designed to show that locutions such as “I seem to see . . .” and “It seems to him as though he hears . . .”—perfectly ordinary, nontechnical locutions—are characteristically used to report on sense-impressions. For convenience, I shall deal explicitly only with

“ . . . seem(s) to see . . .,” but what I shall say about this locution will clearly apply to, e.g., “He seems to hear . . .,” “You seem to feel . . .,” and so forth.

The argument which I have in mind (let us call it *A*) runs as follows:

- (a) Imagine a case in which: (i) Someone, *S*, sees a puddle of blood on his living-room rug; (ii) *S* has reason to believe that he does *not* see blood on the rug; (iii) *S* says “I seem to see blood on the rug.”
- (b) The fact that *S* really does see blood on the rug (though he does not know it) does not make *S* mistaken in saying “I seem to see blood on the rug.” *S*’s utterance makes a true statement, therefore. It follows that *S* is in a certain state in virtue of which what he says is true, as well as being in that state, in virtue of which it is true that he sees blood on the rug.
- (c) But it is impossible to understand what difference between the case under discussion and any other conceivable instance of *S*’s seeing anything would be marked by saying that some instance of *S*’s seeing an object is not an instance in which the state affirmed by “I seem to see . . .” is present. (We can of course conceive of a case which differs from the envisaged case in that *S* does not seem to see the *very same* object which he sees. Thus *S* might seem to see a rat whereas what he truly sees is an old shoe. However, we cannot conceive of a case which differs from the envisaged case in that *S* sees some object at a certain place but fails to seem to see *any object of any sort* at that place.)
- (d) It follows that “I seem to see . . .” affirms the existence of a state which one must, logically, be in if he is to see anything. But “I seem to see . . .” does not mean the same as “I see . . .,” and, hence, it does not affirm the existence of precisely the same state the existence of which is affirmed by “I see” This may be expressed by saying that what “I seem to see . . .” affirms is only a logically necessary, not a sufficient, condition of what “I see . . .” affirms. “I seem to see . . .” is therefore a nontechnical expression which is used to refer to sense-impressions, and Ryle’s claim that there are no such expressions is untenable.

So much for the exposition of *A*. In what follows, I shall be considering and rejecting some criticisms

¹ “Sensation,” *Contemporary British Philosophy*, 3rd Series, H. D. Lewis (ed.) (New York, Macmillan, 1956), p. 435.

of it which some philosophers might wish to employ. I am not at all certain that these criticisms of *A* are the only ones worth considering. But, if my reasons for rejecting them are sound, then this paper may be viewed as, at the very least, an indication that any philosophers wishing to construct a really cogent criticism of *A* would do well to stay clear of the ground which I shall be covering.

II

Ryle says of sentences of the same sort as "I seem to see . . ." that they are used for the purpose of making "guarded statements of what I am tempted or inclined to judge to be the case . . ."² And G. J. Warnock says that "the essential function of the language of "seeming" is that it is noncommittal as to the actual facts. . . ."³ One plausible interpretation of these remarks is as follows. The role of "seems," on all occasions of its use, is exactly similar to the role of "maybe": both words, whenever they are uttered, are used solely for the purpose of asserting a proposition in a guarded or tentative manner or indicating a noncommittal attitude toward the proposition.⁴

A criticism of *A* emerges from this analysis of "I seem to see. . . ." In step (b) of *A* it is said that *S*'s utterance makes a true statement. But, given the similarity of "seem" to "maybe," this will not do. Guarded or noncommittal expressions like *S*'s cannot be used to make false statements and therefore they cannot be used to make true ones. The point may be illustrated by considering a sentence containing "maybe," e.g., "Maybe it will rain today." What is said by this sentence is compatible both with its raining today and with its not raining today and, hence, there is no conceivable state of affairs which can falsify it.⁵ But then, since what is said by "Maybe it will rain today" cannot be false, it cannot be true either. For what cannot be false can be true only if it is a necessarily true proposition like that expressed by, e.g., " $2 + 2 = 4$." And it can hardly be maintained that what is expressed by "Maybe it will rain today" is such a proposition. Exactly similar considerations apply to *S*'s utterance. But now, since *S*'s utterance does not make a true statement, it must be a mistake to say—what is also said in (b)—that "*S* is in a certain

state in virtue of which what he says is true."

Step (d) fares no better in view of the similarity of "seem" to "maybe." It is a gross mistake to claim that "I seem to see . . ." is used to affirm the existence of a state of the speaker which is logically necessary but not sufficient for his seeing some object. The mistake is exactly similar to that which one would be making were he to claim that "Maybe it will rain today" affirms the existence of something which is logically necessary but not sufficient for rain today. The trouble is that "I seem to see *x*" doesn't affirm of the speaker something in any way different from what we affirm of him when we say that he sees *x*, and, hence, it doesn't affirm the existence of something which is only necessary, not sufficient, for his seeing *x*. Just in case, and to the extent that, "I seem to see *x*" is used to affirm anything at all of the speaker, it affirms (in a guarded manner) exactly what we affirm when we say that he sees *x*.

The criticism of *A* just presented is based on the claim that "seem," in all its occurrences (including its occurrences in conjunction with "see"), behaves like "maybe." What may appear to recommend this claim is that, if it were true, then it would be easy to see why "I seem to see *x*" is not generally uttered when the speaker has no doubt that he sees *x*. One does not generally make statements in a guarded manner, nor evince a noncommittal attitude toward propositions, when he is certain of their truth. Moreover, it might appear, at least at first glance, that unless we adopt the analysis of "I seem to see *x*" under discussion, we shall be powerless to explain why "I seem to see *x*" is not uttered when the speaker is convinced that he sees *x*. But this, as I shall subsequently try to show, is mere appearance. And, anyway, there are certain considerations which leave us no choice but to repudiate the analysis now in question and, along with it, the criticism of *A* just considered.

First of all, while it is true that "I seem to see *x*" is not generally uttered by one who firmly believes that he sees *x*, it is plainly false that it cannot be uttered in perfect propriety by one who firmly believes that he does *not* see *x* but suffers an hallucination instead. And this is sufficient to refute the claim that "seem," like "maybe," functions in *all* its occurrences, including its occurrences in

² *Ibid.*, p. 495.

³ *Berkeley* (London, Penguin Books, 1953), p. 186.

⁴ Unless, of course, they are placed in quotation marks and used to refer to word types or tokens.

⁵ It might be argued that the speaker's being sure that it will or will not rain today falsifies what is said. But this is a mistake, since "Maybe it will rain today" does not, strictly speaking, *assert* that the speaker is uncertain about whether it will rain. The speaker, is, of course, being *insincere* (hiding his real belief about the matter) if he is convinced that it will or will not rain

connection with forms of the verb "to see," as a device for enabling the speaker to make a guarded or tentative statement or to evince a noncommittal attitude toward a proposition. For it is less than completely candid for one to assert a proposition tentatively or to evince a noncommittal attitude toward it when one is convinced that the proposition is false, and perceptual propositions are no exception to this rule. *S* cannot say with perfect candor "Maybe I see blood on the rug" when he firmly believes that he suffers an hallucination. But it would be perfectly proper for him to say "I seem to see blood on the rug" in these same circumstances. It follows that, in case the speaker believes that he hallucinates, "I seem to see . . ." does not mean the same as "Maybe I see . . .," i.e., it is not used as a guarded or noncommittal utterance.

Someone may wish to reply that at least the analysis under discussion and the criticism of *A* which is based on it hold good for the case in which *S* is *uncertain* about whether he sees blood on the rug (rather than convinced that he does not). It may be said that, in *this* case, at any rate, *S* would simply be asserting in a guarded way that he sees blood on the rug or evincing a noncommittal attitude toward this proposition. But I think that it is dubious that *S* would here mean something quite different by "I seem to see blood on the rug" from what he would mean were he to be convinced that he suffers an hallucination. Suppose that he starts off in a state of doubt and says "I seem to see blood on the rug" and that he subsequently becomes convinced that he hallucinates and says "I still seem to see the blood." Surely he would not be speaking inappropriately in saying the latter, and yet his use of the word "still" indicates that his second utterance has the same sense as the first one. Moreover, even if we waive this consideration, the reply in question is certainly not sufficient to sustain the proposed criticism of *A*. *Qua* attack on *A*, the reply may be circumvented by the simple expedient of imagining that *S* is convinced that he suffers an hallucination. It will not do to answer that in that case "I seem to see blood" is used by *S* to state that he hallucinates and, hence, is plainly not used to affirm that *S* is in a state which is logically necessary for genuine vision. The claim that *S* uses "I seem to see blood" to state that he is suffering an hallucination entails that *S* says what is false (i.e., since *S* does in fact see blood on the rug). And this last thesis is certainly mistaken. Just as one who says "I seem to see *x*" does not say

something which is falsified by the fact that he suffers an hallucination in which he *merely* seems to see *x*, so too he does not say something which is falsified by the fact that he does *not* suffer an hallucination but really sees *x* instead.

There is another consideration which has a bearing on the present discussion. If it were really the case that "I seem to see . . ." meant the same as "Maybe I see . . .," then the criticism of step (b) of *A* given earlier would be accurate, at least to this extent: There would be no conceivable state of affairs in virtue of which what is said by "I seem to see . . ." could properly be called false—just as there is no conceivable state of affairs in virtue of which what is said by "Maybe it will rain today" can properly be called false. But in fact this is not the case. Though what is said by "I seem to see *x*" is not incompatible either with the speaker seeing *x* or with his suffering an hallucination in which he merely seems to see *x*, it *is* incompatible with his suffering an hallucination in which he seems to see some object other than *x*. One who says, e.g., "I seem to see a brown horse at place *P*" when in fact he suffers an hallucination in which he seems to see a purple dragon at place *P* tells a falsehood. Moreover, what is said by "I seem to see *x* at *P*" is incompatible with the speaker's seeing (really seeing) some object *other* than *x* at *P* which does not *look* like *x* to him. Thus, if I see a cat on the mat and it does not look like a dog to me, I tell a falsehood by saying "I seem to see a dog on the mat." Similar considerations do not hold for "Maybe I see *x*." States of affairs which involve the speaker suffering an hallucination in which he seems to see some object other than *x* do not falsify what he says by this utterance. Though what the speaker seems to see is a purple dragon, he would not be speaking falsely—indeed it might be that he would not be speaking insincerely—were he to say "But maybe I see (really see) a brown horse (instead)." Moreover, the state of affairs in which the speaker sees *y* and it does not look like *x* to him does not falsify (nor necessarily render insincere) what he says when he says "Maybe what I see (really) is *x* (which looks like *y*)."

The essential point here can be put as follows: If I say "I seem to see *x* at *P*" and it is later discovered either that I suffered an hallucination in which what I seemed to see at *P* was not *x* but *y* or that I really saw *y* at *P* and that it did not look like *x* to me, then *eo ipso* I am convicted of having said what is false. If, on the other hand, I say "Maybe I see *x* at *P*" and these same things are

later discovered, then I cannot be accused of having said what is false (nor is it necessarily the case that I can even be accused of being misleading). It follows that "I seem to see . . ." does not mean the same as "Maybe I see. . ."

A final word in this connection. I do not wish to deny that one who says "I seem to see . . ." may be either asserting in a guarded way that he sees the object which he mentions or evincing a noncommittal attitude toward the proposition that he sees it. All that I have tried to establish in this section is that this is never *all* that one does, nor—since "I seem to see . . ." has basically the same sense when the speaker is in doubt as when he is convinced that he is hallucinated—the primary thing that one does, when he says "I seem to see. . ."⁶ If I am right, then the criticism of *A* which I have been considering is without foundation.

III

Another claim about "I seem to see . . ." from which it follows that *A* is incorrect is that "I seem to see . . ." has the same meaning as "I am inclined to believe that I see . . ."⁷ At first glance it may appear that this claim does not really differ from the claim that "I seem to see . . ." means the same as "Maybe I see. . .," since it is tempting to suppose that "I am inclined to believe . . ." plays a role exactly similar to "maybe." But it is, I think, difficult to reconcile the thesis that "I am inclined to believe . . ." has the same use as "maybe" with the fact that "He is inclined to believe . . ." and "You are inclined to believe . . ." are *not* characteristically guarded or noncommittal expressions. I can say of another person that he is inclined to

believe a certain proposition *p* even though I am entirely convinced that *p* is false (or true). When I say "He is (you are) inclined to believe that *p*," I am not, therefore, asserting *p* guardedly nor am I evincing a noncommittal attitude toward it. And there is no other proposition which I might be asserting guardedly or toward which I might be evincing a noncommittal attitude. It follows that the expressions in question characteristically affirm something in a non-tentative way about the person to whom they refer. And it is difficult to believe that this is not also the case when "I" is substituted for "you" and "he" in these expressions. I take it, therefore, that "I am inclined to believe that I see . . ." typically makes an unguarded affirmation of the existence of a certain state of the speaker—the state of being inclined to believe that he sees the object which he mentions.

The analysis now under consideration, unlike the analysis considered in II, cannot be rejected on the grounds that there are no conceivable states of affairs the existence of which would falsify what is said by "I am inclined to believe that I see *x* at place *P*." States of affairs in which the speaker is not inclined to believe that he sees *x* at *P* but is inclined to believe instead that he sees some object other than *x* at *P* would falsify what is said by the utterance in question. And it is not at least immediately obvious that such states of affairs do not coincide with those states of affairs, mentioned in II, in virtue of which what is said by "I seem to see *x* at *P*" would be false. Moreover, some philosophers have argued that it is possible for one to be inclined to believe that a certain proposition is true, and, at the same time, for him to disbelieve that very proposition.⁸ If these philosophers are

⁶ It may be objected that Ryle and Warnock possibly did not intend to adopt the strong position I have been attacking but simply the weaker one mentioned above. But if Warnock would admit that it is not the sole or primary function of "seem," in all its occurrences, to enable the speaker to evince a noncommittal attitude toward a proposition, then I think that (in the absence of a supplementary analysis on Warnock's part) he would be admitting in effect that his argument against Berkeley's idealism (*Berkeley*, pp. 181–189) is without force. (Unfortunately it would take us too far afield were I to try to substantiate this here.) As for Ryle, he appears to believe that it follows immediately from the claim that "seem" is always used to make guarded or tentative or noncommittal statements that "I seem to see . . ." does not report the occurrence of sense-impressions. It is not at all plain that this follows, however, if "I seem to see . . ." is *more* than a guarded assertion that the speaker sees an object or a device for evincing a noncommittal attitude toward this proposition.

⁷ D. M. Armstrong says "To say 'It looks oval to me' usually means that I have some inclination to believe that I am seeing something oval." (*Perception and the Physical World* [London, Routledge & Kegan Paul, 1961], p. 92.) No doubt Armstrong would be willing to extend this type of analysis to "I seem to see. . ."

⁸ See, for example, Alan R. White, "The Causal Theory of Perception," *Proceedings of the Aristotelian Society, Supplementary Volume XXXV* (1961), p. 165. White appears to wish to defend the claim that locutions like "I seem to see *x*" mean the same as "I am inclined to think that *x* has such and such visual characteristics." (*Ibid.*, p. 167.) The claim is false, however. A man blind from birth can be inclined to think that *x* has such and such visual characteristics but cannot truthfully say that he seems to see *x*. And similar considerations apply to another analysis of "I seem to see *x*" which White appears to find adequate, viz., "I am inclined to think that what is before my eyes is *x*." (*Ibid.*, p. 167.)

It should be pointed out that the analysis being considered above is not subject to this same criticism. A man blind from birth can truthfully say neither that he seems to see *x* nor that he is inclined to believe that he sees it.

right, then the present analysis of "I seem to see. . .," unlike the analysis considered in II, is not at least plainly incompatible with the fact that I can say "I seem to see x " with perfect propriety even though I disbelieve that I see x .

A criticism of A which may be based on the present analysis is as follows. Though step (b) of A is legitimate (it is perfectly correct to say that "I am inclined to believe that I see blood on the rug" makes a true statement and does so in virtue of a certain state of S ; hence, the same holds for "I seem to see blood on the rug"), steps (c) and (d) are plainly false. We can easily conceive of a case which differs from the envisaged case in that, though S sees some object at P , he is not in the state affirmed by "I seem to see. . ." For this is simply to conceive of a case in which S is not just inclined to believe that he sees some object at P but has no doubt whatever that he does so.

One approach for the philosopher who wishes to defend A vis-à-vis this criticism is to deny that a person may both be inclined to believe that something is the case and, at the same time, convinced that it is not. For in case this claim is false, then the present analysis of "I seem to see. . ." is, like the analysis in II, incompatible with the fact that it is appropriate for one to say "I seem to see. . ." when he is convinced that he suffers an hallucination. This much, at least, can be said on behalf of the thesis that disbelieving a proposition is incompatible with being inclined to believe it: One who told us that he was inclined to believe that some proposition was true when in fact he was convinced that it was false would ordinarily seriously mislead us by so doing. To this it may be replied: (a) that, while being inclined to believe p may be incompatible with being *fully convinced* (firmly believing) that not- p , it is not incompatible with believing not- p with some lesser degree of conviction,⁹ and (b) that people who suffer hallucinations are not (and indeed cannot be) fully convinced that they do not see the object which they seem to see. But while (a) is unexceptionable, (b) appears to be false. It appears that the concepts of hallucination and of being fully convinced of something are not such that it is absurd to say of someone that he is fully convinced that he suffers an hallucination. Indeed, it is true as a matter of fact that hallucinogenic drugs produce hallucinations which the subject knows full well to be hallucinations.

But let us waive this rebuttal. Perhaps it is the case that one's being inclined to believe that he

sees x is compatible with his being absolutely certain that he does not see it. There are two further, and, I think, stronger arguments against our accepting the analysis of "I seem to see x " in terms of an inclination on the speaker's part to believe that he sees x . (1) We do not ordinarily say "He seems to see x " or "You seem to see x " when we are convinced that the person to whom we refer really does see x . Just as one who says "I seem to see x " generally indicates that he doubts or disbelieves that he really sees x , so one who utters the second and third person forms of "seem to see" generally indicates that he, the speaker, doubts or disbelieves that the person to whom he refers really sees x . But, unfortunately for the proponent of the present analysis, similar considerations do not hold for "He is inclined to believe that he sees x " and "You are inclined to believe that you see x ." We frequently say of other people (though not of ourselves) that *they* are inclined to believe some proposition about the truth of which we, the speakers, have absolutely no doubt. (2) I can believe, or be inclined to believe, that I see (really see) an old shoe on my table when it looks to me as though I see (I seem to see) a rat there instead. That is to say, I can believe, or be inclined to believe, that what *seems* to me to be a rat is in reality an old shoe. I cannot, however, seem to see an old shoe on my table when it looks to me as though I see (I seem to see) a rat there instead. Being inclined to believe that one sees an old shoe cannot, therefore, be the same as seeming to see an old shoe. And of course we may generalize: being inclined to believe that one sees x cannot be the same as seeming to see x .

IV

The analysis of "I seem to see. . ." just rejected, like the analysis considered in II, would, if we could accept it, enable us to give an account of the fact that we do not ordinarily say that we seem to see some object when we have no doubt that we do see the object. Whatever may be thought of the claim that a person may be convinced that a certain proposition is *false* and yet inclined to believe it, it is indisputable that one may not be both convinced that a certain proposition is *true* and also inclined to believe it. It would plainly be absurd to say "I am inclined to believe that I see blood on the rug and I have no doubt at all that I do." Exactly similar considerations would, of course, be true of seeming to see some object if (what is not the case) this

⁹ Possibly this is White's position in the paper cited in the preceding footnote.

were the same as being inclined to believe that one sees the object.

Another claim which would, if true, entail that *A* is incorrect and account for the fact that we do not say that we seem to see *x* when we have no doubt that we see *x* is the claim that "I seem to see . . ." means the same as "I doubt that I see . . ." If this were so, then (1) "I seem to see *x* and I have no doubt that I do" would be self-contradictory and (2) steps (c) and (d) of *A* would be obviously wrong. But this claim, like the analysis considered in II, cannot accurately represent the meaning of "I seem to see . . ." in those cases in which the speaker is convinced that he suffers an hallucination. If he is at all concerned to be accurate, one does not say that he doubts that something is the case when he is convinced that it is not the case. It follows that it is unlikely that the present analysis is an adequate analysis of the meaning of "I seem to see . . ." on any occasion of its utterance. Moreover, the present analysis can be rejected on grounds similar to those set out in III. "He doubts that he sees . . ." and "You doubt that you see . . ." do not indicate doubt or disbelief on the part of the speaker, while "He seems to see . . ." and "You seem to see . . ." do indicate doubt or disbelief. Also, though one can doubt that he sees *y* at *P* when he seems to see *x* at *P*, one cannot seem to see *y* at *P* when he seems to see *x* at *P*.

There is still another claim from which it follows that *A* is incorrect and which would, if true, explain why we do not say "I seem to see *x*" when convinced that we do see *x*. I have in mind the claim that "I seem to see . . ." means the same as "I disbelieve that I see . . ." If this analysis were correct, then, once again, "I seem to see *x* and I have no doubt that I do" would be self-contradictory. Moreover, steps (c) and (d) of *A* would be clearly mistaken. But the analysis may be rejected on the grounds: (1) that it cannot account for the fact that people say "I seem to see *x*" when they strongly suspect that they may in reality be seeing *x*; (2) that "He disbelieves that he sees *x*" and "You disbelieve that you see *x*" do not indicate that the speaker doubts or disbelieves that the person to whom he refers sees *x*, while this is not true of "He seems to see *x*," etc.; (3) that one can disbelieve that he sees *y* at *P* when he seems to see *x* at *P*, but cannot seem to see *y* at *P* when he seems to see *x* at *P*.

At this point, it may appear that I have saved *A* only at the cost of losing any hope of finding an explanation of the fact that we do not ordinarily

say that we seem to see an object unless there is at least some doubt that we do. But in fact a very simple and plausible explanation of this fact remains open. If *A* has been successfully vindicated, then we must grant that one who says "I see . . ." tells us in part that he seems to see some object. But now, in at least most circumstances in which one believes firmly that he does see the object which he seems to see, it would be pointless for him to withhold the additional information (whatever it may be) conveyed by "I see. . . ." One who believes that he sees a certain object and says only that he seems to see it, deliberately withholds information from us—tells us part, but only part, of what he believes to be the case. And ordinarily there is simply no reason for one to do this. Of course, one may not be motivated to say anything at all; we do not often find it worth while reporting that we see the objects which in fact we see. But when one is motivated to report that he sees a certain object, he is generally motivated to tell us what he believes to be the whole story (that he sees the object) and not just a part of it (that he seems to see the object). It is for this reason that one who says "I seem to see . . ." is ordinarily understood not to be fully convinced that he sees the object which he mentions.

An analogy may be useful at this point. If I wish to report that I went for a walk this afternoon, I will not, at least ordinarily, say "I moved my legs this afternoon." To say this latter is to tell only part of the story, while to say "I went for a walk" is to tell the part told by "I moved my legs" and more. And ordinarily there would be no reason for me to wish to tell only the part of the story about my moving my legs even though I know that the whole story is true. It is for this reason that, were I to say "I moved my legs this afternoon," I would generally be understood to be indicating that my ability to walk had been in some way impaired (an analogue to the fact that one who says "I seem to see . . ." is generally understood to doubt or disbelieve that he really sees the object to which he refers).

It goes without saying that, if the foregoing explanation is correct, then it is not self-contradictory or in any way absurd to say that we seem to see objects whenever we see them. Indeed, my explanation embodies a repudiation of the thesis that it is self-contradictory or absurd to say this, since it involves the claim that seeming to see is a logically necessary condition of seeing. Philosophers who are convinced that "I see . . ." says

something which is in some sense logically incompatible with what is said by "I seem to see . . ." will of course reject the explanation which I have offered (along with *A*). And perhaps they will be right in so doing. But at least they cannot argue for their position on the grounds that any of the analyses of "I seem to see . . ." which have been

considered in this paper are correct. Nor can they argue for their position by saying that unless *some* such analysis were adequate it would be impossible to present an even speciously credible account of the fact that "I seem to see *x*" is not generally uttered except when the speaker doubts or disbelieves that he really sees *x*.

Vanderbilt University

VII. CAN A SMELL OR A TASTE OR A TOUCH BE BEAUTIFUL?

FRANCIS J. COLEMAN

ALMOST all aestheticians exclude smell, taste, and touch from their definitions of beauty. These senses are passed over as though it were a thing admitted on all hands that they could never be beautiful. Even the aestheticians who do believe that some argument for excluding them is called for, generally appeal to linguistic usage: one does not *speak* of a beautiful odor or a beautiful taste. And that is certainly true. We should all be a little taken back if the person across from us at table sighed "Beautiful!" after he had savoured a certain wine or smelled a certain dish. But for the sake of argument, let us suppose that just such a thing happens. Let us suppose that we have been served grenadine of beef, that our companion smells it for a few moments, and then says, "It is a beautiful odor, isn't it?" How should one go about convincing him that an odor—and by extension a taste or a touch—is not the sort of thing that can be beautiful? What arguments should one draw up?¹

Before proceeding, however, it may be well to observe how the answer to the question that I have

asked will affect certain other questions often raised in aesthetics. First, let us consider the aesthetic experience. Whatever else that term may involve, it is clear that if we deny the possibility of beautiful smells, tastes, and tactual sensations, then an aesthetic experience of a country landscape, for example, would rest entirely upon its visual and auditory properties. But if we grant that certain smells can be beautiful, then they could sometimes be elements of the aesthetic experience as well. Second, whether we give an objective, subjective, or even exclamatory analysis of the aesthetic judgment, it is clear that if we deny the possibility of beautiful smells and tactual sensations, then it would be inappropriate to refer to such data or aggregates of such data in an aesthetic judgment. Third, if we give "beauty" a formal as opposed to an expressive definition, as Kant for example does, then the reasons that a person could give for calling an object "beautiful" would never involve a reference to the senses of smell, taste, or touch. For example, in explaining why one thought Keats's "St. Agnes Eve" to be beautiful, one should

¹ Most philosophers who have dealt with this question have denied the possibility that non-visual and non-auditory sense data can be beautiful. Kant, in the *Critique of Judgment* maintains that one must not call an object beautiful if it merely pleases us. (Cf. *op. cit.*, tr. J. H. Bernard [New York, 1951], p. 47). Santayana, in *The Sense of Beauty* (New York, Charles Scribner's Sons, 1936, pp. 51, 52, 53), maintains that tastes "have never been so accurately or universally classified or distinguished," and that the art of cooking "deals with a material far too unrepresentable to be called beautiful." Denis Diderot concludes that tastes and smells are not beautiful because one does not call them beautiful. After having defined "beautiful" as "tout ce qui contient en soi de quoi réveiller dans mon entendement l'idée de rapports," he says, "Quand je dis tout, j'en excepte pourtant les qualités relatives au goût et à l'odorat; quoique ces qualités pussent réveiller en nous l'idée de rapports, on n'appelle point beaux les objets en qui elles résident . . ." (*Oeuvres complètes*, Tome dixième [Paris, Garnier Frères, 1876], p. 26). St. Augustine holds a middle position. When speaking of the attractions of smells and sounds (*inlecebra odororum* and *voluptates aurium*) he seems to grant that such things can be beautiful, but one ought not to succumb to them. (*Confessions*, Société d'Éditions "Les Belles Lettres," [Paris, 1947, vol. 2], p. 276). Prall also concedes that smells, tastes, and "vital feelings" can be the materials of beauty. "We cannot rule out the specific character of tastes and of bodily feelings and of smells from the materials of genuine aesthetic experience on any clear ground." But since "smells and odors do not in themselves fall into any known or felt natural order or arrangement, nor are their variations defined in and by such an intrinsic natural structure, as the variations in color and sound and shape give rise to in our minds," they are only "elementary aesthetic materials." (*Aesthetic Judgment* [New York, Crowell, 1929], pp. 57, 75.) I am acquainted with only two aestheticians who hold without reservation that "beautiful" can be used of all sense data. Guyau, in *Les problèmes de l'esthétique* (Paris, Ancienne Librairie Germer Baillière et Cie., 1902, p. 61) writes: "Pour nous, nous croyons que toute sensation agréable, quelle qu'elle soit, et lorsqu'elle n'est pas par sa nature même liée à des associations répugnantes, peut revêtir un caractère esthétique en acquérant un certain degré d'intensité, de retentissement dans la conscience." Thomas Munro is of much the same opinion; in speaking of the "lower-sense arts," he writes: "According to the definition of art recommended here, products and services appealing aesthetically to the lower senses are included." (*The Arts and Their Interrelations* [New York, Liberal Arts Press, 1947], p. 136.)

not make mention of its extremely sensuous imagery. To use Kant's words, such images may add to the "charm" but not the "beauty" of the poem.² We see, then, that the question that we are considering is not isolated from others frequently asked in aesthetics.

Surely we would not contend that there is a logical contradiction in "beautiful odor" as there is in "square circle." For let us adopt the most common definition of "beautiful," that of the Oxford English Dictionary: "Full of beauty; pleasing to the senses or intellect." It could apply to a smell or a tactual sensation as readily as to the other senses.

Our first step, then, in trying to convince our opponent that he ought not to speak of beautiful odors and tastes would be to invoke English usage. "It is just not done," we would say: "It is a linguistic impropriety." But his retort would be only too just: many reasonable things are not done and many proprieties are ridiculous. "What I need," he would say, "is a *reason* for not speaking of a beautiful odor or taste. I do not need to be reminded that English-speaking people generally avoid such expressions. We do not generally speak of gold mountains and virtuous horses; but we can conceive what such things would be. Think of the Houyhnhnms. So, too, if I should not speak of beautiful odors or tactual sensations—except as a joke or hypothetically—it should be because no such things exist."

Let us attempt to convince him by another argument. "Odors and tastes and tactual sensations cannot be beautiful because a simple datum is never beautiful." "And by a 'simple datum' you mean?" he would ask us. "I mean a patch of solid color or a single chord." "Well and good," he would reply. "But still, are not single colors beautiful to some persons? Mr. Ducasse finds violet beautiful.³ For my part burnt sienna is beautiful. And certain chords, like the one that links the Second and Third Movements of Beethoven's 'Appassionata Sonata', are beautiful."

Although we would grant him his point, we would try to explain it away. Even though people do speak of particular colors and sounds as beautiful, it is only because the context of that color or sound has been omitted or temporarily ignored. It is always the context—the relations of the chord with others in the musical composition, or the relations of the color with others in the painting—

that renders the particular color or chord beautiful. A color is beautiful *for* something else, for a person's complexion, for a chair in a room decorated in a certain way, for a certain area in a painting. A color considered singly is never beautiful.

Our opponent would fall silent: dogmatism often has that effect. He would recall how persons often find a particular color beautiful, just in itself, without relating it to anything else. But let us suppose that our opponent passes over this. "Still," he could say, "even if a simple sense datum is never beautiful, not all tastes are simple and many odors and tactual sensations are of considerable complexity. Consider the touch of silk, which, by the way, I find beautiful. It is smooth, not like marble, but soft and minutely textured and supple; it is also a dry touch, not like paper, but cool and almost glassy. Even though we feel all these qualities of silk at the same instant, still upon a little reflection, we can abstract the various sensations that go together to give us the sensation of silk."

Let us grant him that some tactual sensations are complex; perhaps we would also grant that some smells are complex. Lily-of-the-valley and hyacinths are similar in that both have a heavy sweet odor; but the former has a tincture of wild grass and the latter a certain bitterness. "But even so," we would argue, "sounds and colors are naturally so arranged that one may be more beautiful because of its combinations with others. A deep olive green, for example, would be more beautiful if it were set against a tannish pink. But when we apprehend various flavors and smells, we have nothing but a mere succession of experiences, each of which is distinct, none of which heightens or diminishes the other. This lack of order, this uniqueness, this inability to combine with others so that an organic whole is formed, is the reason why smell, taste, and touch should not be spoken of as beautiful."

"But think of certain culinary dishes in which ingredients are combined to form a whole. The mind may wander over them, discern the different flavors, and note how they complement and contrast with each other. A well-prepared dish is an 'organic whole' in basically the same sense as a well-conceived painting: elements are contrasted with each other but not contrasted so sharply that they fail to cohere. We often find that a dish 'needs something,' that 'there is something missing,'

² Immanuel Kant, *Critique of Judgment*, tr. J. H. Bernard (New York, 1951), p. 58.

³ C. F. Ducasse, "Esthetic Contemplation and Sense Pleasure: A Reply," *Journal of Philosophy*, vol. 40 (1943), pp. 156, 159.

just as we sometimes do with a painting. This is not to say that as far as their aesthetic merit or worth are concerned a well-prepared dish and a well-conceived painting are the same, but only that both are fit subjects of an aesthetic judgment."

"But," we would object, "the criteria of appropriate combinations of tastes is simply a question of local customs and disposition."

And our opponent would retort that much the same thing is true of the criteria of combining colors and sounds. Consider the greatly different ways in which Eastern and Western peoples organize tones. Or, consider what colors the natives of New Guinea put together; for us such combinations are jarring and repugnant.

"Yes," we would interrupt. "Be that as it may. Nonetheless, beauty has to do with form. Although odor, taste, and touch may add to the pleasurable-ness that we derive from an object, those senses contribute nothing to its beauty." In our eagerness to convince him, we might refer to the article "Esthetic Contemplation and Its Distinction from Sense Pleasure":⁴ "The contemplation of a flower may be heightened by its fragrance. . . . But while this is so, and we may assume that the experiences are esthetic, the contribution which odor makes is, I submit, not esthetic."

But our opponent would refer to the fact that most persons would find a rose with an odor more beautiful than one without. Or that if a person's tactual sense were deadened by an anaesthetic, the beauty of velvet or silk would almost cease to exist for him.

But we would overlook this and cite the proof that Mr. Zink gives in the same article: "The evidence is that the particular fragrance qualifying a particular flower, for example, either might be exchanged for another perfume, or even removed entirely, without altering the visual composition of the object. Thus color and shape constitute the nucleus of esthetic form, to which odors attach. Could we not discover equal esthetic worth in the rose and the violet if their perfumes were transposed?"⁵

"But what you give as proof is not proof at all!" our opponent would exclaim. "Of course the odor of a flower might be exchanged or removed altogether without altering the visual composition of the object. Odors are not seen. You might as

well argue that an actor's voice does not contribute to his theatrical effectiveness because if his voice were removed, his visual appearance would not be altered. Furthermore, though it is true that we could discover equal aesthetic worth in the rose and the violet if their fragrances were transposed, for it is only by habit that we associate a particular odor with a particular object, nevertheless, that the odors could be transposed does not prove that an odor does not contribute to the total aesthetic character of an object."

At this point we would be certain that our opponent was confounding aesthetic appreciation with sense pleasure, and continue quoting from the same article: "There is another sense, however, in which the perception of relation—and in visual art, of spatial relations—is the essence of esthetic appreciation. Perception of the relations *internal to* the object, recognition of the location of one part or element in relation to the others, is the vital force in appreciation of art. This is the trait that distinguishes esthetic appreciation from sense pleasure."⁶ "In sense pleasure that satisfaction is more complex the more concentrated is attention upon the simple pleasurable quality. . . . In esthetic appreciation, on the other hand, it is necessary that thought perform the positive function of specific and minute examination of the complex art-object to which esthetic pleasure fastens."⁷

But our opponent would speak up for the great mass of mankind who have no technical knowledge of music and still find some pieces beautiful; he would speak up for those who do not see the visual relations in painting, but are sometimes aesthetically moved; he would speak up for those who could not engage in a "specific and minute examination" of a poem, but feel "aesthetic pleasure" nonetheless. Are such persons "not really" having an aesthetic experience?

Furthermore, what does it mean to speak of "the relations *internal to* the object?" Let us briefly consider that expression. First, it would imply that there are at least two elements in the object—unless aesthetic appreciation could also consist in perceiving the relation of identity, which is absurd. Second, in what ways could the elements be "internally related"? An element could touch another, either spatially or temporally. An element

⁴ Sidney Zink, "Esthetic Appreciation and Its Distinction from Sense Pleasure," *Journal of Philosophy*, vol. 39 (1942), p. 709.

⁵ *Ibid.*

⁶ *Ibid.*

⁷ *Ibid.*, p. 705.

could follow another, either spatially or temporally. Or an element could resemble another or be unlike it. Or elements could vary in degrees of quality or quantity. And lastly, an element could be the cause of another. In the arts, this last relation would be rarely applicable. But in representative painting, for example, one could speak of cause and effect if the subject matter of the painting was so related. But since we perceive all of these relations among elements which are non-beautiful and positively not beautiful, it is difficult to conceive how "the perception of relations" could be "the essence of esthetic appreciation."

We would interrupt and say that it was not so much these relations which we had in mind. We meant the notions of harmony and disharmony, balance and unbalance, proportion and disproportion. It is these that one must perceive in order to have an aesthetic experience.

But does one *perceive* balance? Does one *perceive* discord? Is it not more accurate to say that we observe differences in size and shape, degrees of colors, degrees of loudness and intensity, and that some of these please us and others jar us? Two persons could readily agree upon the relative intensity of the colors of "The Man with the Glove," but one find them harmonious and the other discordant. Two persons could readily agree that one of the bell towers of Chartres Cathedral was taller than the other, and that their tops were quite different. But they could measure the towers forever and not discover whether they were in harmony with the rest of the edifice or with each other. The "esthetic relations" of balance and unbalance, harmony and disharmony, proportion and disproportion, or whatever others might be set forth, are no more than instances of the more general relations that all things in the world may be subject to, but qualified with our commendation or censure, our approval or disapproval, our delight or abhorrence. And since odors and tastes are subject to some of these relations, for the elements of a complex taste may vary in quality and degree, they cannot be shut out from beauty on the ground that the mind can perceive no relations among them.

By this point our patience with our dining companion would naturally be exhausted. We would seize about for any sort of argument to convince him. "But smells and tastes are so completely utilitarian that they do not lend themselves to that disinterestedness, that contemplative detachment which colors and sounds are capable of."

Our opponent would ask us to consider how much more useful sight and hearing are to us than taste and smell, and how much more difficult it is to get along without them. Yet in a cool hour, when there is nothing pressing in upon us, we can give ourselves up to a painting or a piece of music. We can look at something merely to be looking at it. So, too, with the senses of smell and taste. Though we often use these senses merely as tools, sometimes we can taste or smell something for its own sake.

We would seize about for another argument. "The reactions of mankind to smells and tastes vary so greatly, not only one man's with another's, but at different hours of the day in the life of one and the same man, that it would be impossible to speak of them as beautiful."

Yet there is considerable unanimity in the tastes of mankind. Consider how whole nations can be in accord over certain foods, and how they come almost to sanctify them with the term, "national dish." Consider how readily makers of perfume rely upon the uniformity of persons' likes and dislikes. But let one exaggerate the variety of human reactions as much as one will, they would reveal fewer vicissitudes than do their reactions to works of art. There was a time when Shakespeare was generally condemned and disliked, when Gothic architecture was called hideous and oppressive, and when Beethoven's works were even denied to be music.

"But surely you are forgetting that there is no scale, no intelligible order in which odors or tastes may be arranged, as there are for sounds and colors. Given any tone, we can determine what its third or fifth or seventh must be. Given any color, we can place it according to its intensity. Surely such things are impossible with smells and tastes."

"Indeed, they are impossible. But how does it follow that tastes and smells are therefore not beautiful?"

"Because we have shown something common to all things called 'beautiful,' that is, that their auditory or visual properties can be ordered, which is not the case . . ."

"But consider some of Rathko's paintings in which only two colors are set one against the other. The beauty of such paintings rests upon our seeing that the one color complements or contrasts with the other. This same thing sometimes occurs when two odors or two tastes are compounded, even though there are no "scales" of odors or tastes."

We would grope about for another argument, but suddenly realize that there was none left. But there always remains abuse: "Smells, tastes—how could they be beautiful? They are nothing but the *lower senses*. And let us even assume that some of them could be beautiful, still can you conceive what deleterious consequences it would have to give ourselves up to enjoying a touch or an odor for its own sake? Decadence, moral dissoluteness . . ."

Let us leave these two men while they are still friends. We have listened to the arguments. We have seen that none provides a good reason for excluding smells, tastes, and tactual sensations from the things that can be beautiful. Let us now cast about for some good reasons for including them. But let me note that I am not interested in altering English usage; I could not even if I were. Even physicists continue to speak of sunrise and sunset. I should only like to make it clear that from the point of view of aesthetics there are good reasons for including data from all our senses among the things that can be beautiful.

First, let us consider whether a blind man infers or directly experiences the beauty of a face or a statue that he touches. Let him be a man blind from birth so that there is no chance that he merely recollects having seen the object and that he once found it beautiful. Of the few blind persons I have known, none has ever hesitated in pronouncing upon the beauty of objects that they were acquainted with by touch. They say, "Yes, it is beautiful," but not, "I suppose it is beautiful," or "It must be beautiful for those that can see it." What they say is surely a good indication that they are not inferring that the object would be beautiful *if* they could see or that it would be beautiful for those who can see. Nevertheless, some doubt could be cast upon this verbal evidence, for persons sometimes use the wrong words and locutions. However, surely one would know whether one is inferring or directly experiencing. One knows, for example, that one infers the bent stick in water really to be straight. One does not see that the stick is bent. So, too, one could tell whether one directly experiences the beauty of an object, or infers that it would be beautiful for us if we could see it, or that it would be beautiful for those who can see it. But blind persons would reject

such a description of what they are doing when they call a statue or a face "beautiful." They say that they are not inferring, and that by calling a certain object "beautiful," they mean it is so to them, not just to those who could see it. Doubt cannot be cast upon this as evidence: they know the English language, they have no reason for lying. I conclude, then, that touch, or more exactly, the sensations given to us through the sense of touch and organized by us, can be beautiful. This is not to say that the statue is not more beautiful for one who can see it, or perhaps one should say, beautiful in a different way. But the beauty of the Elgin Marbles, for example, is in an important sense accessible to the blind.

Arguments parallel to this cannot be given for the other fine arts because all the rest can be perceived by only one sense.⁸ Music is heard; painting is seen. But poetry has the peculiarity of being perceived by none of the senses and by all of them: its effectiveness depends upon our memories derived from all the senses, and these memories are revived in us through words. Now if we canceled or replaced the words that refer to data derived from the senses of smell, touch, and taste, a poem would be less effective. If "To Autumn" were raped of such words as "mellow,"⁹ "sweet," "warm," "clammy," "fume," and others, and were replaced with expressions referring to data derived solely from sight and hearing, it would be less beautiful. Although it does not follow from this observation that smell, touch, and taste must therefore be capable of being beautiful when considered individually, it does follow that reference to them augments the beauty of a literary composition.

Second, let us consider what it entails to speak of the beauty of velvet, for example. Although it would be naive to exclude some reference to its costliness and uncommonness, we cannot ignore the warmth, delicate thickness, and suppleness of its touch. These traits contribute to its beauty as much as its luster and iridescence.

Third, let us glance at the expression, "a beautiful dinner." Although this expression would refer to many aspects of the dinner—the quality of the silver, china, and linen, the promptitude and unobtrusiveness of the service—still one must include the taste, texture, and aroma of the various

⁸ There are of course the composite arts of the drama, the dance, and the opera, needless to be mentioned. It is also true that one can sometimes feel the vibrations of music; perhaps the totally deaf could appreciate music in this way.

⁹ I include "mellow" because part of its meaning in the poem is derived from our tactual sensations—the softness of ripe fruit.

dishes, and their harmony one with another.

Fourth, let us consider the behavior of a person enjoying a short piece of music and again while enjoying the odor of a flower. We shall find much the same quiet and receptive resignedness, the same sweet tranquillity of expression.

But, it will be objected, considering the mere behavioral signs is not sufficient. Aesthetic contemplation is highly intellectual and demanding; enjoying an odor or a taste is sensuous and passive.

To answer this objection, let us overlook the difficulty of distinguishing "intellectual" from "sensual." Rather, let us notice that some art makes no demands upon us at all. Consider once again some of Rathko's paintings in which two colors are set against each other at opposite ends of the panel. Nothing is demanded of us; there is no "specific and minute examination" that we must perform to appreciate the work. Or consider the Japanese haiku. Again, we have no intellectual task to perform. We have only to read it through, which we do at a glance. There are no questions to be asked. We either "see" it or do not. Consider this example:

Spring night:
Blowing his flute,
A passerby.¹⁰

It is of course true that by attending only to a person's behavior one cannot be certain whether he is enjoying a poem; consequently, one cannot infer that enjoying an odor, for example, can be an aesthetic experience because it closely resembles the phenomenal behavior of a person enjoying a short lyric or piece of music. What I mean to point out is that to distinguish the two experiences by calling the one non-intellectual and the other intellectual is sometimes inappropriate. To use such words as "thinking" and "understanding" to describe our experience of certain lyrics and melodies is as inappropriate as to use such words to describe our experience of odors and tactual sensations. There is of course intellectual and demanding art. The poetry of Milton, Donne, and Eliot must be thought through in order to be

appreciated. Much of Bach and Prokofieff are of the same sort. But there are also the lyrics of Sappho, some songs of Mozart, and some colored blots of contemporary French artists that demand nothing from us. They approach us, linger with us, and abandon us like the odor of peonies.

But let us consider even a most complex work of art, one that must be studied and analyzed to be appreciated aesthetically. Still, there is something else of comparable complexity, but not a work of art at all: the aesthetic delight of being physically with a person one loves. All the senses are swept into activity; the delight is held on to for its own sake.

I conclude then that in some instances sense pleasure is a species of aesthetic pleasure. When the data from our senses of smell, taste, and touch are attended to for their own sake, when we entertain them not to learn something from them, or to predict something on the basis of them, or merely to satisfy our wants with them, and when the data are of a certain intensity, however short-lived they may be, then they can be beautiful. However, aesthetic delight in one's senses is not to be confounded with sensual pleasure; for the latter entails a satisfaction of our appetites or basic animal needs. To take aesthetic delight in one's senses, one's needs must already be in part satisfied, though not to the point of satiety.

But as I write these words I am afraid that I have not convinced anyone who was not of my opinion to begin with, for at heart the disagreement is one of attitude. There are those who hold that certain of our senses are base and not to be yielded to; there are those who believe them to be morally neutral. There are those who believe that beauty must be lofty and intellectual; there are those who add that it can also be sensuous, thoughtless, and simple. There are those who hold sexual intercourse to be solely for procreation; there are those who think that it can be the occasion of aesthetic delight. But here, driven as I am to the opposing camps of Christianity and Paganism, I must leave off, for it is a disagreement that I could hardly hope to resolve.

University of Pittsburgh

¹⁰ *Anthologie de la poésie japonaise* (Paris, Librairie Orientaliste Paul Geuthner, 1935), p. 173. My translation.

VIII. CORRIGENDA

Volume I (1964)

BRIAN MEDLIN: *The Unexpected Examination* (pp. 66–72).

Page 68, col. 2, last line: Add to sentence . . .the information (I): $p_1 \vee p_2 \vee p_3$.

K. W. RANKIN: *Referential Identifiers* (pp. 233–243).

Page 233, title: Replace “Indentifiers” by “Identifiers”

HERBERT SPIEGELBERG: *Towards a Phenomenology of Experience* (pp. 325–332).

Page 327, col. 1, thesis 3.4: Delete the word “no”

Page 329, col. 2, line 8: Delete comma after “expectations”

Page 330, col. 2, line 12 from bottom: Replace “sequences” by “sequence”